



Universidad Central de Venezuela  
Facultad de Ciencias  
Escuela de Computación  
Centro de Ingeniería de Software y Sistemas  
Laboratorio de Inteligencia Artificial

**Aplicación de la minería de datos  
para la clasificación de la morfología  
de espermatozoides humanos**

Trabajo Especial de Grado  
presentado ante la ilustre  
Universidad Central de Venezuela  
por el Bachiller  
Alexander José Martínez Zambrano  
para optar al título de  
Licenciado en Computación

Tutores:  
Profa. Haydemar Nuñez  
Prof. Roberto Casañas

Septiembre 2009.

## AGRADECIMIENTO

En primer lugar, agradezco a Dios, por guiarme y permitirme culminar este proyecto con éxito.

Agradezco a mis tutores por brindarme su apoyo, conocimiento y guía para alcanzar el cumplimiento de las metas.

Agradezco a mis familiares y amigos por todo el apoyo brindando.

## RESUMEN

El análisis de líquido seminal es un examen de laboratorio que evalúa las diversas propiedades del semen. Una de las características que determina este examen es la morfología espermática, en la cual se calcula el porcentaje de espermatozoides con y sin alteraciones morfológicas.

Esta investigación plantea el desarrollo de una aplicación que, mediante el uso de algoritmos de minería de datos, ejecute el proceso de clasificación morfológica espermática de imágenes de muestras de semen.

Esta aplicación fue implantada usando el lenguaje de programación JAVA y permite realizar la carga de imágenes digitales de las muestras de semen, segmentar y extraer valores de dichas imágenes y clasificar los espermatozoides utilizando los valores obtenidos.

Palabras claves: Análisis de Líquido Seminal, análisis morfológico, clasificación espermática, clasificación de imágenes, minería de datos.

# CONTENIDO

## MARCO TEÓRICO

---

AGRADECIMIENTO.....	2
RESUMEN.....	3
CONTENIDO.....	4
INTRODUCCION.....	6

## **MARCO TEÓRICO.....8**

### **ANÁLISIS DE LÍQUIDO SEMINAL HUMANO.....8**

1.1 Análisis de líquido seminal	8
1.1.1 Estudio Macroscópico	8
1.1.2 Estudio Microscópico	9
1.2 Morfología espermática	11
1.3 Resultados del análisis de líquido seminal	13

### **MINERÍA DE DATOS.....16**

2.1 ¿Qué es la minería de datos?	16
2.2 Proceso de minería de datos	17
2.2.1 Limpieza de los datos	17
2.2.2 Preprocesamiento	18
2.2.3 Extracción de Conocimiento	18
2.2.4 Interpretación y Evaluación	19
2.3 Tareas de la minería de datos	19
2.3.1 Tareas predictivas	19
2.3.2 Tareas descriptivas	20
2.4 Técnicas de minería de datos	20
2.4.1 Redes neuronales	20
2.4.2 Árboles de decisión	24
2.4.3 K-Vecinos	26
2.4.4 Reglas de Clasificación	27

### **AUTOMATIZACIÓN DEL ANÁLISIS MORFOLÓGICO ESPERMÁTICO.....29**

3.1 Planteamiento del problema	29
3.2 Antecedentes	30
Análisis de semen de cerdo.....	30
Clasificación intracelular del espermatozoide.....	31
Segmentación automática de espermatozoides.....	32
3.3 Propuesta de investigación	32
3.4.1 Objetivo general	33
3.4 Objetivos	33
3.4.2 Objetivos específicos	33
3.5 Estimación del modelo de clasificación	33
3.5.1 Descripción de la Base de Datos	34
3.5.2 Limpieza y preprocesamiento de los datos	36
3.5.3 El problema de las clases desbalanceadas	38
3.5.4 Experimentos y resultados	40
3.6 Desarrollo de la aplicación	44
3.6.1 Análisis de requerimientos	45
Requerimientos funcionales	45
Requerimientos no funcionales	45
3.6.2 Modelo de Casos de uso	45
3.6.3 Arquitectura de la aplicación	48
3.6.4 Plataforma tecnológica	50
3.6.5 Interfaz de la aplicación	50
3.6.6 Verificación de la aplicación	53
<b>CONCLUSIONES.....</b>	<b>56</b>
Aportes	58
Recomendaciones y futuros trabajos	58
REFERENCIAS.....	59

# INTRODUCCION

El análisis de líquido seminal es un examen de laboratorio que evalúa las diversas propiedades del semen, con la finalidad de detectar posibles problemas de esterilidad. Éste consta de varias pruebas, en las cuales se analizan diversos aspectos de la muestra de semen, entre los que destaca el análisis de la morfología espermática, el cual determina el porcentaje de espermatozoides con y sin alteraciones morfológicas.

La prueba de la morfología espermática se realiza en un microscopio, observando diferentes zonas de la muestra para determinar la cantidad de espermatozoides normales y anormales; pero debido al entorno de trabajo, los equipos utilizados para el estudio y el personal que lo realiza, los resultados de este examen varían en menor o mayor grado.

Esta investigación plantea el desarrollo de un sistema automatizado que permita realizar el proceso de clasificación de la morfología espermática a partir de imágenes de muestras de semen, empleando algoritmos de minería de datos.

El presente documento se encuentra estructurado en dos secciones: *Marco teórico*, el cual muestra información asociada con la construcción del sistema automatizado; y el *Marco aplicativo*, que contiene información acerca del desarrollo del sistema automatizado.

El marco teórico está distribuido en varios capítulos:

- **Capítulo 1:** En el cual se describe el proceso de análisis de líquido seminal humano.

- **Capítulo 2:** Contiene conceptos referentes a la minería de datos, que fueron empleados en el desarrollo del sistema automatizado de clasificación.

En cuanto al marco aplicativo se contempla lo siguiente:

- **Capítulo 3:** Describe el proceso de minería de datos utilizado para determinar el modelo de clasificación. Además, se presentan los aspectos relacionados con la construcción del sistema automatizado: definición de los requerimientos, diseño de las funcionalidades, realización de pruebas y otros aspectos relevantes del proceso de construcción del sistema automatizado.

Finalmente, se presentan las conclusiones y las referencias en las cuales se basó el trabajo.



### Análisis de líquido seminal humano

*El análisis del semen es utilizado para medir la cantidad y calidad del semen y de los espermatozoides de un hombre.*

#### 1.1 Análisis de líquido seminal

El análisis de líquido seminal es un examen de laboratorio que permite diagnosticar la calidad del semen; su finalidad es evaluar sus propiedades y detectar posibles problemas de esterilidad [VAS07]. Este estudio se divide en dos partes, un estudio Macroscópico y uno Microscópico.

##### 1.1.1 Estudio Macroscópico

Es el estudio donde se evalúan las características del líquido seminal, en éste se determina el aspecto, volumen, viscosidad y Ph de la muestra [OMS01]. A continuación se describen los elementos que se evalúan en esta etapa:

1. **Aspecto:** Es la apariencia del líquido seminal. Una muestra normal es homogénea y de color gris opalescente. Cuando la muestra es opaca,

indica baja cantidad de espermatozoides; y si presenta color marrón indica que contiene glóbulos rojos.

2. **Volumen:** Es el espacio físico ocupado por el líquido seminal. El volumen debe ser mayor de 2ml. Si una muestra presenta menos de 2ml se denomina hipospermia y podría indicar una obstrucción causada por una infección genital, alteración congénita de los vasos deferentes o eyaculación retardada.

3. **Viscosidad:** Se refiere a la consistencia del semen, es decir, la resistencia de un líquido al movimiento o flujo. Una viscosidad disminuida se observa cuando la gota del semen cae con rapidez y una viscosidad aumentada cuando la muestra presenta un filamento con una longitud mayor a 1cm.

4. **Ph:** Índice que expresa el grado de acidez o alcalinidad de una solución. Se evalúa distribuyendo una gota de semen sobre una tira de papel ph con un rango entre dos y nueve. Dentro de los primeros 30 segundos el color debe ser uniforme y se compara con la plantilla de calibración. El valor de referencia es de 7.2.

### 1.1.2 Estudio Microscópico

Es el estudio donde se evalúan las características de los espermatozoides. Consta de un conjunto de pruebas para determinar la concentración, movilidad, vitalidad, aglutinación y la morfología de los mismos [OMS01].

1. **Concentración de espermatozoides:** Es la cantidad de espermatozoides presentes en la muestra. La concentración espermática se evalúa en un hemocitómetro (porta objeto del microscopio), realizando un conteo de espermatozoides en varios sectores de la muestra; el

número de sectores pueden ser cinco, 10 o 25 cuadros. A partir del número de espermatozoides de esos cuadros, se aplica un factor de corrección y finalmente se obtiene la concentración expresada en millones de espermatozoides por milímetro de semen.

2. **Movilidad:** Es la determinación del número y porcentaje de espermatozoides que se mueven. Se evalúan 200 espermatozoides como mínimo de acuerdo a las siguientes categorías:

- a. Movilidad progresiva rápida.
- b. Movilidad progresiva lenta.
- c. Movilidad no progresiva.
- d. Inmóviles.

3. **Vitalidad:** Es la determinación de la cantidad y porcentaje de espermatozoides vivos presentes en el eyaculado.

4. **Aglutinación:** Es la determinación de la cantidad y porcentaje de espermatozoides unidos entre ellos por cualquiera de sus partes, ya sea cabeza con cabeza, pieza intermedia con pieza intermedia, cola con cola o cualquier otra combinación. La unión de espermatozoides inmóviles entre ellos o de espermatozoides móviles con restos celulares u otro tipo de células no debe considerarse aglutinación.

5. **Morfología espermática:** Es el análisis de la forma y estructura de los espermatozoides, es uno de los principales análisis de una muestra seminal y para éste se toman en cuenta las tres partes del espermatozoide.

## 1.2 Morfología espermática

El análisis de la morfología espermática consiste en clasificar a los espermatozoides en dos categorías, normales y anormales. Según las normas de la Organización Mundial de la Salud [OMS01], se consideran espermatozoides normales a todos aquellos que cumplan con las siguientes características:

- Cabeza ovalada, única, lisa, acrosoma ocupando entre 40% y 70%, 4-5 $\mu$ m de longitud, 2.5-3.5 $\mu$ m de ancho.
- Pieza intermedia estrecha, unida axialmente, alineada con la cabeza, sin restos citoplasmáticos, que no debe exceder de un 1/3 del área de la cabeza, longitud 6-10 $\mu$ m
- Cola uniforme, única, desenrollada, 45 $\mu$ m de longitud, ligeramente más estrecha que la pieza intermedia, sin residuos citoplasmáticos.

Cualquier espermatozoide que no cumpla con lo anterior es considerado anormal. En la Figura 1 se puede observar las diferentes anomalías que puede presentar un espermatozoide.

El protocolo que sigue un experto para el conteo morfológico de los espermatozoides es el siguiente [OMS01]:

- Examinar la preparación con objetivo de inmersión (100x).
- La evaluación morfológica se lleva a cabo en varias áreas del portaobjetos, en las cuales no hayan grandes aglomeraciones de espermatozoides.

- Contar 200 espermatozoides, expresando el porcentaje de normales y anormales, estos últimos clasificados en las categorías de defecto de cabeza, pieza intermedia y cola.

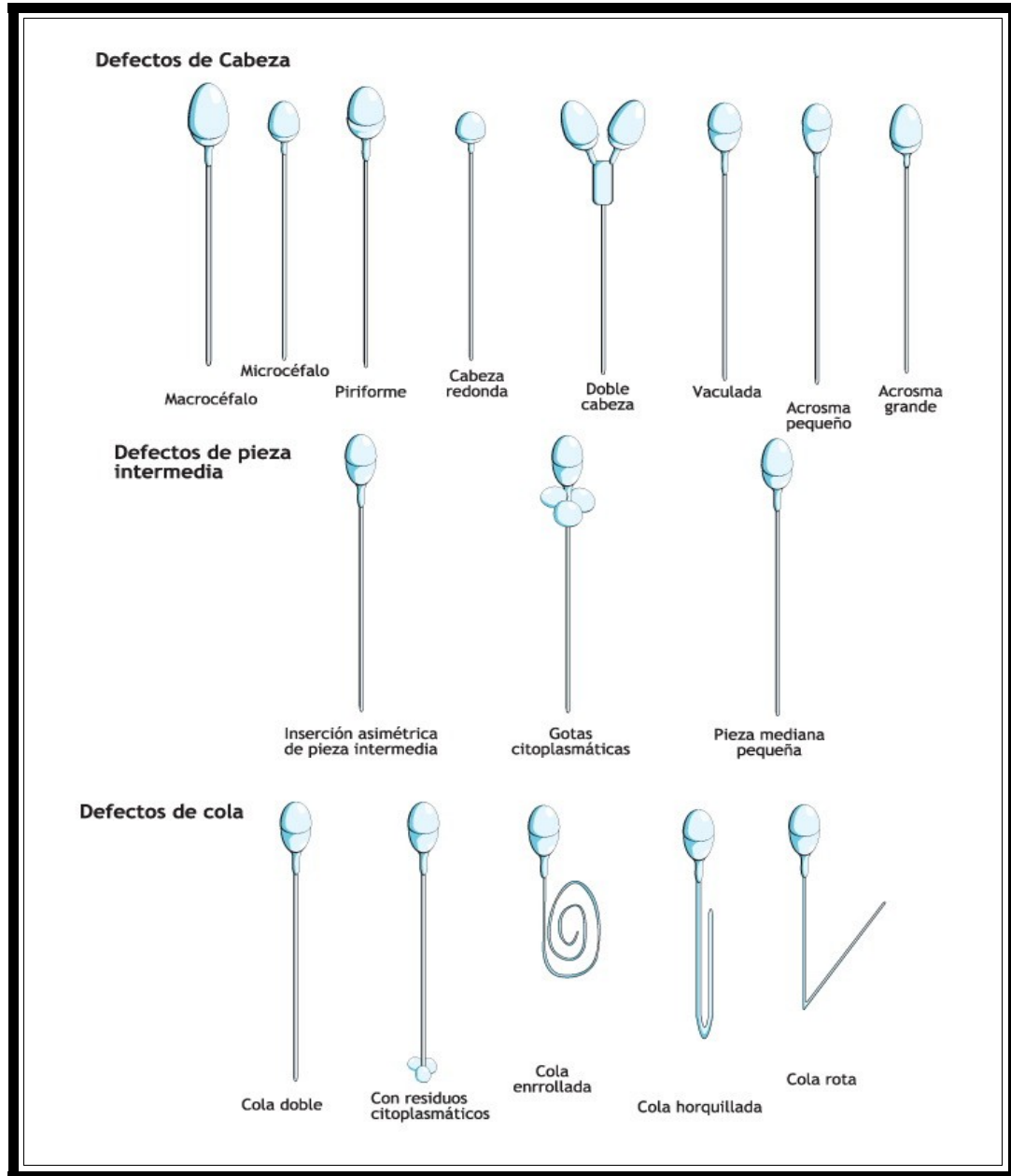


Figura 1: Anormalidades del espermatozoide

- Se recomienda contar por duplicado, utilizando diferentes preparaciones de la misma muestra. Esto permite reducir la variabilidad y los errores de contaje, este procedimiento es especialmente valioso en los casos en que el diagnóstico y tratamiento depende de la morfología espermática.
- Sólo los espermatozoides se consideran reconocibles para el estudio morfológico, las células inmaduras no se cuentan. Las cabezas aisladas o colas aisladas no se cuentan, pero deben ser reportadas separadamente.
- Los espermatozoides morfológicamente anormales, frecuentemente tienen defectos múltiples. Se debe registrar el número de defectos divididos entre el número de espermatozoides anormales, lo cual representa el índice de teratozoospermia, o índice de anomalías múltiples (TZI). El índice de teratozoospermia debe estar comprendido entre 1.0 (cada espermatozoide anormal tiene un único defecto) y 3.0 (cada espermatozoide anormal tiene defectos de cabeza, pieza intermedia y cola).

### **1.3 Resultados del análisis de líquido seminal**

De acuerdo a los criterios de la OMS [OMS01] los valores normales de un análisis deberían ser los siguientes:

- Volumen: 2-5ml.
- Color: Blanquecino.
- pH: 7.2
- Viscosidad: Normal
- Motilidad: Mayor a un 50%
- Concentración: Mayor a 20.000.000 por ml
- Morfología: Al menos el 30% de los espermatozoides deben tener una morfología normal.

Cuando los valores de los parámetros medidos en la muestra no cumplen con las referencias de la OMS puede ocurrir alguna de las siguientes anomalías:

1. **Hiperespermia:** Ocurre cuando el volumen del eyaculado es superior a 6ml [VAL02].
2. **Hipospermia:** Ocurre cuando el volumen del eyaculado es inferior a 2ml [VAL02]. Existen 3 niveles de Hipospermia:
  1. **Hipospermia leve:** Volumen mayor a 1.5ml y menor a 2ml.
  2. **Hipospermia moderada:** Volumen entre 1ml y 1.5ml.
  3. **Hipospermia severa:** Volumen menor a 1ml.
3. **Azoospermia:** Es la ausencia total de espermatozoides en el semen [OMS01]. Existen dos tipos de azoospermia:
  1. **Azoospermia obstructiva**, donde el espermatozoide es creado, pero no puede mezclarse con el resto de los fluidos eyaculatorios debido a una obstrucción física.
  2. **Azoospermia no obstructiva**, donde hay un problema de espermatogénesis.
4. **Aspermia:** Ausencia de eyaculado externo [OMS01].
5. **Astenozoospermia:** Se presenta cuando menos del 50% de espermatozoides tienen una progresión lineal [OMS01]. Existen 3 niveles de Astenozoospermia:
  1. **Astenozoospermia leve:** Porcentaje de espermatozoides con progresión lineal entre 30%-50%.
  2. **Astenozoospermia moderada:** Porcentaje de espermatozoides con progresión lineal entre 20%-30%.

3. **Astenozoospermia severa:** Porcentaje de espermatozoides con progresión lineal menor a 20%.
  
6. **Oligozoospermia:** Poca cantidad de espermatozoides en el líquido seminal, su concentración es menor a 20.000.000 por ml [OMS01]. Existen distintos niveles de Oligonospermia:
  1. **Oligozoospermia leve:** Cantidad de espermatozoides es mayor a 15.000.000 por ml y menor a 20.000.000 por ml.
  2. **Oligozoospermia moderada:** Cantidad de espermatozoides es mayor a 10.000.000 por ml y menor a 15.000.000 por ml.
  3. **Oligozoospermia severa:** Cantidad de espermatozoides es menor a 10.000.000 por ml.
  
7. **Teratozoospermia:** Menos del 30% de espermatozoides con morfología normal [OMS01].
  
8. **Oligoastenoteratozoospermia:** Se utiliza cuando el semen eyaculado presenta bajo recuento, baja motilidad y morfología con más del 70% de formas anómalas [OMS01].
  
9. **Necrozoospermia:** Es cuando más del 50% de los espermatozoides se están vivos pero inmóviles [CHA03]. Existen 3 niveles de Necrozoospermia:
  1. **Necrozoospermia leve:** Ocurre cuando el porcentaje de espermatozoides vivos inmóviles está entre 50%-60%.
  2. **Necrozoospermia moderada:** Ocurre cuando el porcentaje de espermatozoides vivos inmóviles está entre 60%-70%.
  3. **Necrozoospermia severa:** Ocurre cuando el porcentaje de espermatozoides vivos inmóviles es mayor al 70%.



10. **Normozoospermia:** La muestra cumple los parámetros de normalidad establecidos por la OMS [OMS01].

## Minería de datos

*La minería de datos prepara, sondea y explora los datos para obtener la información que puede estar oculta en ellos; utilizando técnicas de inteligencia artificial y el análisis estadístico para abordar soluciones a problemas de predicción, clasificación y segmentación.*

### 2.1 ¿Qué es la minería de datos?

La minería de datos se define como la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de un conjunto de datos [GAR08]. Es un paradigma utilizado en los programas de gestión del conocimiento como soporte a la toma de decisiones. Sus bases se encuentran en la inteligencia artificial y en el análisis estadístico. Mediante los modelos extraídos utilizando técnicas de minería de datos se aborda la solución a problemas de predicción, clasificación y segmentación.

## 2.2 Proceso de minería de datos

El proceso para la realización de un proyecto de minería de datos es siempre el mismo (Figura 2), independientemente de la técnica específica de extracción de conocimiento usada [VAL06]:

1. **Limpieza de los datos.**
2. **Preprocesamiento.**
3. **Extracción de Conocimientos.**
4. **Interpretación y Evaluación.**

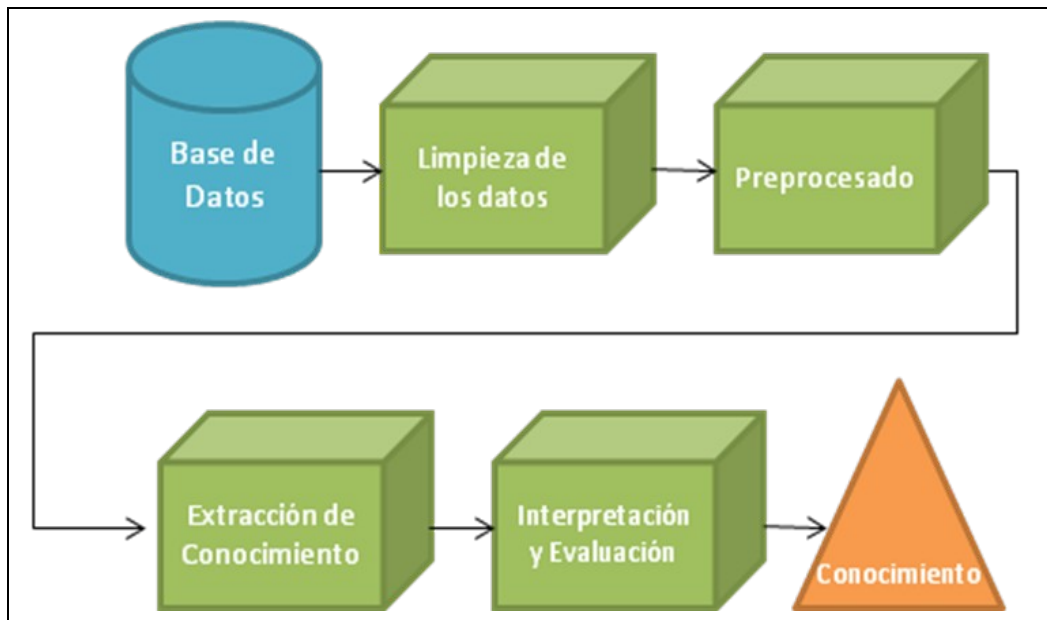


Figura 2: Proceso de Minería de datos

### 2.2.1 Limpieza de los datos

El formato de la información contenida en la fuente de datos por lo general no es el más adecuado y la mayoría de las veces no es posible utilizar ningún algoritmo de minería. Mediante la limpieza de datos se filtran los

datos, de forma que se eliminan valores incorrectos, no válidos y/o desconocidos; según las necesidades y el algoritmo que va a utilizarse, se obtienen muestras de los mismos (en busca de una mayor velocidad de respuesta del proceso), o se reduce el número de posibles valores.

### 2.2.2 Preprocesamiento

Aún después de haber limpiado los datos, existe una gran cantidad de información que debe ser preprocesada antes de aplicar algún algoritmo de minería de datos. Con el preprocesamiento se busca eliminar atributos, crear nuevos atributos en caso de ser necesario y seleccionar los atributos más influyentes en el problema.

1. **Eliminación de atributos:** Se busca eliminar atributos que no están relacionados con el dominio del problema.
2. **Creación de nuevos atributos:** Permite la creación de atributos a partir de otros atributos, con el objetivo de obtener información que podría ser útil.
3. **Selección de atributos:** Se realiza la evaluación de los atributos mediante algoritmos estadísticos para seleccionar a aquellos que sean útiles al dominio del problema.

### 2.2.3 Extracción de Conocimiento

Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesamiento diferente de los datos.

### 2.2.4 Interpretación y Evaluación

Una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe referirse a alguno de los pasos anteriores para generar nuevos modelos.

## 2.3 Tareas de la minería de datos

El proceso de minería de datos consiste en la generación de modelos o en la mejora de patrones a partir de los datos. Este ajuste normalmente es estadístico, en el sentido que se permite un cierto ruido o error dentro del modelo. Los algoritmos de minería de datos realizan en general tareas de predicción de datos desconocidos y de descripción de patrones [ACO04].

Las tareas predictivas emplean variables o campos de la base de datos para predecir valores futuros de otras variables de interés; mientras que las tareas descriptivas buscan obtener información sobre las posibles relaciones entre los datos.

### 2.3.1 Tareas predictivas

1. **Clasificación:** Es la asignación de una o varias clases predefinidas a cada registro. La meta es inducir un modelo para poder predecir una clase dados los valores de los atributos.

2. **Regresión:** La meta es inducir un modelo para poder predecir un valor numérico dados los valores de los atributos.

### 2.3.2 Tareas descriptivas

1. **Agrupación:** Es la búsqueda de grupos naturales basándose en el principio de maximizar la similitud entre elementos de un grupo minimizando la similitud entre grupos [HER05].
2. **Análisis de asociación:** Es la búsqueda de relaciones entre registros individuales o grupos de registros de la base de datos [QUI07].
3. **Detección de desviaciones, casos extremos o anomalías:** El objetivo es encontrar aquellos elementos o características que son significativamente diferentes del resto de los datos [QUI07].

## 2.4 Técnicas de minería de datos

Existen muchas técnicas de minería de datos, las cuales se diferencian principalmente por el tipo de patrón que generan. Entre éstas se pueden mencionar:

### 2.4.1 Redes neuronales

Es una técnica de aprendizaje y procesamiento automático basado en la forma en que funciona el sistema nervioso a través de modelos matemáticos.

El sistema nervioso está compuesto de neuronas. Una neurona es una célula que consta de un cuerpo semiesférico, del que sale una rama principal, el axón y varias ramas más cortas llamadas dendritas. En la figura 3 se puede observar la estructura de una neurona.

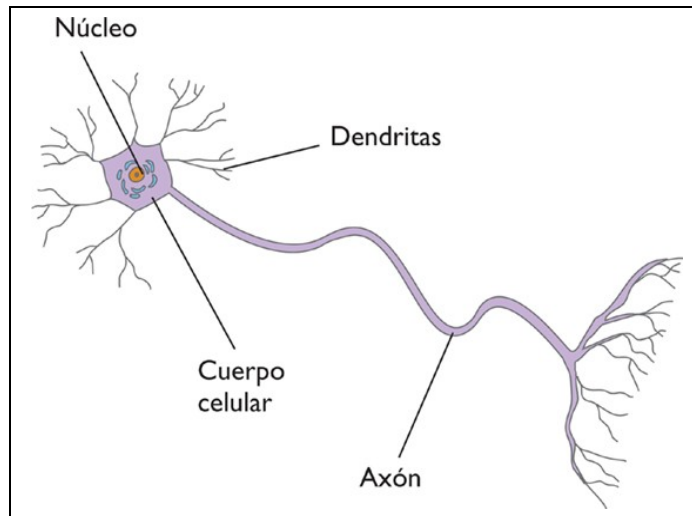


Figura 3: Neurona Biológica

Una de las características de las neuronas es su capacidad de comunicarse o sinapsis. Para esto las dendritas reciben señales de entrada; el cuerpo celular las combina e integra y emite señales de salida. El axón transmite dichas señales a los terminales axónicos, que a su vez distribuyen la información a un nuevo conjunto de neuronas.

Una red neuronal, al igual que el sistema nervioso, se compone de unidades llamadas neuronas, donde cada neurona está conformada por de una red de conexiones de entrada, una función de propagación, una función de activación y una salida por donde se transmite el valor de activación a otras unidades [FLO08]. En la figura 4 se puede ver el modelo de una red neuronal artificial.

- **Entrada:** Es el conjunto de valores de entradas de la neurona, estos pueden ser los valores iniciales a evaluar o valores de salida de otra neurona. Estas entradas son modificadas por pesos que determinan la fuerza de la entrada.

- **Función de propagación o de red:** Calcula el valor base de la neurona, es decir, un valor numérico calculado a partir de las entradas; generalmente es una suma ponderada.
- **Función de activación:** Se encarga de obtener el resultado final de la red.
- **Conexiones ponderadas:** Es el peso que tiene cada valor de entrada de la neurona.
- **Salida:** Es el valor de salida de la neurona obtenido de la función de de activación de la misma. Es la suma de las entradas ponderadas, si la suma es igual o mayor que un umbral definido, entonces la neurona se activa. Esta es una situación de todo o nada; cada neurona se activa o no se activa.

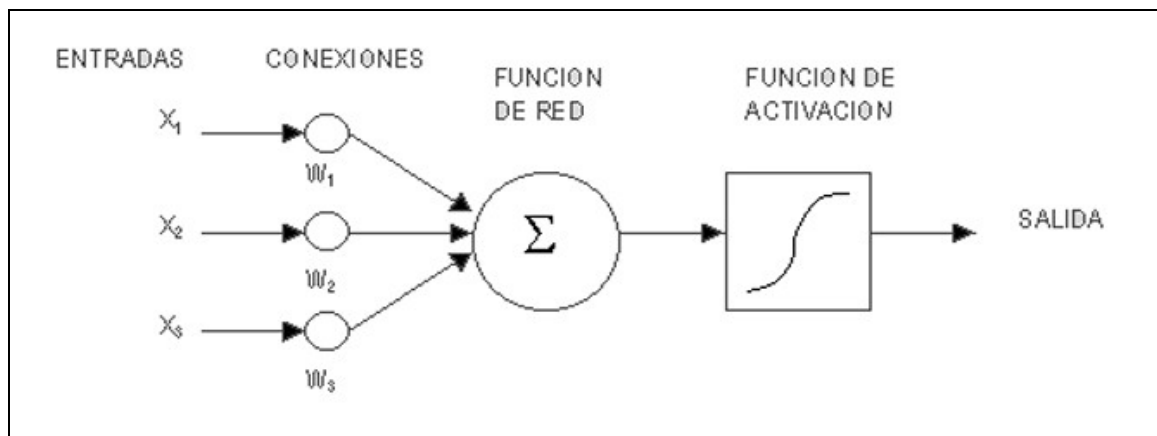


Figura 4: Modelo de una neurona artificial

Es posible clasificar las redes entre redes de una sola capa o monocapa y las redes con múltiples capas.

- **Redes monocapa:** establecen conexiones laterales entre las neuronas que pertenecen a la única capa de la red.
- **Redes Multicapa:** disponen las neuronas agrupadas en varias capas. Normalmente, todas las neuronas de una capa reciben señales de

entrada de otra capa anterior, más cercana a la entrada de la red, y envían su señal de salida a una capa posterior, más cercana a la salida de la red. A estas conexiones se les denominan conexiones hacia delante o feedforward. Sin embargo en un gran número de estas redes también existe la posibilidad de conectar las salidas de las neuronas de capas posteriores a las entradas de capas anteriores, a estas conexiones se les denomina conexiones hacia atrás o feedback.

El proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada se denomina mecanismo de aprendizaje. Los cambios que se producen durante el proceso de aprendizaje se reducen a destrucción, modificación y creación de conexiones entre las neuronas. La creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero; una conexión se destruye cuando su peso pasa a ser cero.

Un aspecto importante respecto al aprendizaje es conocer cómo se modifican los valores de los pesos; cuáles son los criterios para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información, estos criterios determinan la regla de aprendizaje. Se suelen considerar dos tipos de reglas, las que responden a lo que se conoce como aprendizaje supervisado y aprendizaje no supervisado. La diferencia fundamental entre ambos tipos es la existencia o no de un agente externo que controle el proceso de aprendizaje.

- **Redes con aprendizaje supervisado:** El proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor comprueba la salida de la red y en caso de que ésta no coincida con la deseada, se procederá a

modificar los pesos de las conexiones, con el fin de que la salida obtenida se aproxime a la deseada.

- **Redes con aprendizaje no supervisado:** Las redes con dicho aprendizaje no requieren de influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta de una entrada es o no correcta. Estas redes deben encontrar las características, regularidades, correlaciones o categorías que se pueden establecer entre los datos que se presentan en su entrada.

### 2.4.2 Árboles de decisión

Un árbol es una estructura de datos compuesta por un conjunto de elementos, conocidos como nodos, que se encuentran conectados entre sí. Un nodo es un registro de un conjunto de valores que posee referencias a otros nodos y es referenciado por otro nodo. Se dice que un nodo *a* es padre de un nodo *b* si existe un enlace desde *a* hasta *b*; existe un único nodo sin padres, denominado raíz; un nodo que no tiene hijos se conoce como hoja [APT97].

Un árbol de decisión es un modelo de predicción representado por un árbol, en donde:

- Un nodo contiene un test para la evaluación de un atributo y nodos hijos para cada posible respuesta del test.
- Una hoja contiene el nombre de una clase que representa uno de los posibles resultados del árbol.

Los árboles de decisión proveen una visión gráfica de la toma de decisión necesaria, especifican las variables que son evaluadas, qué acciones deben ser tomadas y el orden en el cual la toma de decisión será efectuada (ver

Figura 5). Cada vez que se ejecuta un árbol de decisión, sólo un camino será seguido dependiendo del valor actual de las variables evaluadas [APT97].

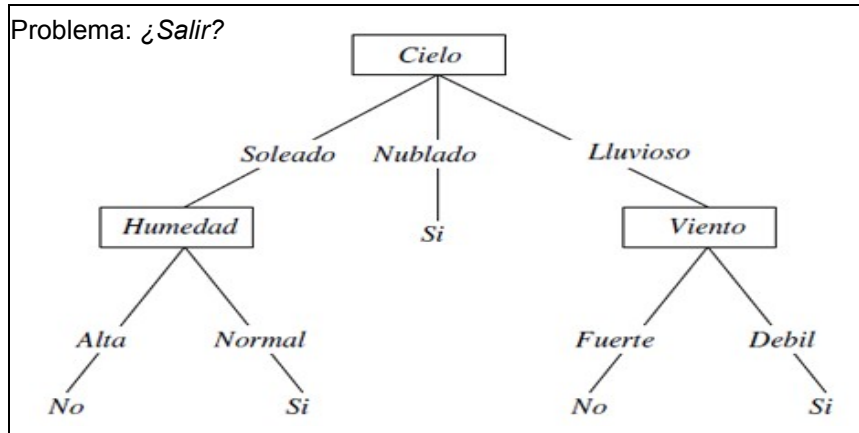


Figura 5: Ejemplo de un Árbol de decisión.

Crear un árbol de decisión es un proceso recursivo sobre los atributos de los datos, a lo cuales se les realiza un cálculo estadístico para medir la ganancia de información, el atributo con las más alta ganancia de información será el atributo más útil para la clasificación.

La ganancia de información está basada en la idea de la entropía, la cual mide la cantidad de información en el atributo, ésta se define como:

$$E(s) = \sum -p(I)\log_2 p(I)$$

Donde “S” es al atributo a evaluar y “p(I)” la distribución de las clases.

Si todos los ejemplos de S pertenecen a la misma clase I, la entropía es 0. Si el valor de la entropía es 1 indica que el resultado del atributo es aleatorio. Con la entropía calculada, el siguiente paso es calcular la ganancia de información:

$$GAIN(S,A) = E(S) - \sum (|S_v|/S) * E(S)$$

La construcción comienza con la determinación del a raíz del árbol, para lo cual se usa el atributo con la mejor ganancia de información, luego este proceso se repite con los atributos restantes por cada rama que se cree.

### **2.4.3 K-Vecinos**

K-Vecinos más cercanos es uno de los más simples algoritmos de aprendizaje automático. En éste, un objeto es clasificado por la mayoría de votos de sus vecinos, con el fin de ser asignado a la clase más común entre sus K-Vecinos más cercanos, en donde “K” es un entero positivo, que se determina mediante un proceso de prueba y error que viene determinado por el problema [PLA07] (ver Figura 6).

Para obtener los vecinos más cercanos de un objeto se calcula la distancia entre éste y todas las instancias vecinas, utilizando atributos continuos. La distancia se determina mediante el cálculo de una función matemática, entre las que destacan la distancia Euclídea y la distancia Manhattan.

El resultado de la clasificación puede ser tanto discreto como continuo. En el caso de una clasificación discreta, el resultado es la clase más común de los K-Vecinos. Para clasificaciones continuas, se puede tomar la media de las clasificaciones.

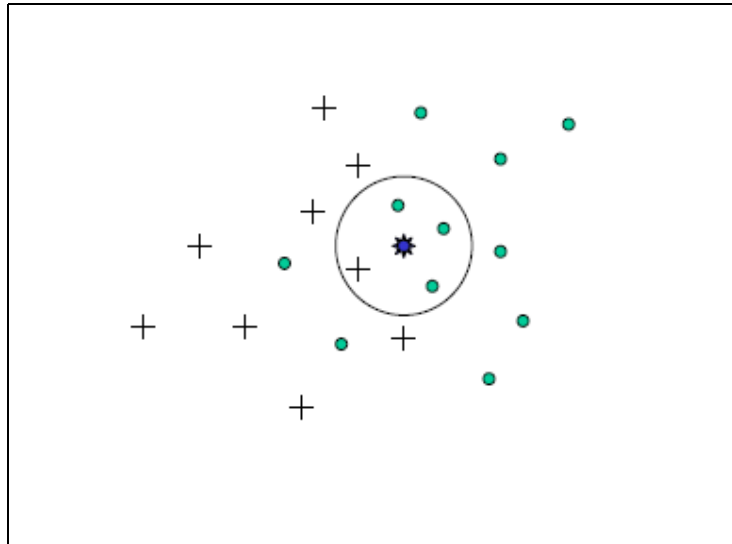


Figura 6: Ejemplo gráfico de una instancia con sus K-Vecinos más cercanos.

A continuación se presenta el algoritmo pseudo formal del algoritmo de K-Vecinos más cercanos:

INICIO

- Entrada:  $D = (x_1, x_2, \dots, x_n)$ , Registros Existentes.
- $Y$  = Nuevo caso a clasificar.
- PARA todo objeto ya clasificado en  $D$
- calcular  $d_i = \text{distancia}(x_i, Y)$ .
- Ordenar  $d_i$  en orden ascendente.
- $D_k = K$  primeros casos.
- Asignar a  $Y$  la clase más frecuente de  $D_k$

FIN

## 2.4.4 Reglas de Clasificación

## MARCO TEÓRICO

---

Una regla es una proposición lógica que relaciona una serie de objetos, ésta incluye dos partes: la condición y la conclusión. La estructura de una regla es:

**IF** <condición> **THEN** <conclusión>

La condición consiste en una expresión lógica con una o más premisas conectadas mediante operadores lógicos “y”, “o” ó “no”; mientras que la conclusión es el resultado obtenido si la condición llega a cumplirse [JOS09]. Un ejemplo de un conjunto de reglas se muestra en la Figura 7.

Las reglas de clasificación son métodos de formalización de datos, usadas para mejorar la toma de decisiones. Pueden ser usadas tanto en procesos de negocio como en investigaciones científicas [APT97]. Existen varios algoritmos para obtener un conjunto de reglas a partir de datos, entre los cuales se pueden citar PRISM y RIPPER [ORA04].

```
R1: IF tiene fiebre THEN padece gripe
R2: IF tiene fiebre THEN padece bronquitis
R3: IF tiene fiebre THEN padece tuberculosis
R4: IF tose mucho THEN padece gripe
R5: IF tose mucho THEN padece bronquitis
R6: IF tose mucho THEN padece tuberculosis
R7: IF tiene dolores musculares THEN padece gripe
R8: IF tiene dolores musculares THEN padece bronquitis
R9: IF tiene dolores musculares THEN padece tuberculosis
R10: IF tiene buen aspecto THEN padece gripe
R11: IF tiene buen aspecto THEN padece bronquitis
R12: IF tiene buen aspecto THEN padece tuberculosis
```

Figura 7: Ejemplo de reglas de clasificación.



## Automatización del análisis morfológico espermático

*La necesidad de automatizar el proceso de análisis de semen humano, con el fin de brindar efectividad a los resultados del mismo, es la motivación que ha guiado este trabajo investigativo.*

### **3.1 Planteamiento del problema**

El análisis de líquido seminal es un estudio que presenta una alta complejidad para su realización, éste incluye diversas pruebas, una de las cuales es la Morfología Espermática, el cual basa sus resultados en el análisis visual que realiza el experto a la muestra de semen.

La morfología espermática es un proceso que consta de diversos pasos y cuyo objetivo es analizar la estructura del espermatozoide a partir de las muestras de semen. El proceso se realiza usando un microscopio y debido a los factores variables que pueden estar presentes al momento de evaluar la muestra, además de la subjetividad visual del experto, dos estudios de la misma pueden generar resultados diferentes. Debido a esto sería conveniente disponer de herramientas que automaticen este proceso, como una forma de apoyar a los especialistas que realizan este tipo de examen.

En este sentido, se han desarrollado algunos trabajos que buscan esta automatización, como los que se presentan a continuación.

### **3.2 Antecedentes**

Automatizar los procesos de la vida diaria ha sido la meta que se quiere alcanzar en muchos aspectos, con el fin de llegar a disminuir errores causados por la intervención humana en procesos que requieren un alto grado de exactitud. Se han realizado diversos prototipos de segmentación automática de cuerpos biológicos utilizando técnicas de procesamiento de imágenes para analizar muestras, que anteriormente se debían estudiar directamente desde el microscopio por un experto. Dichas técnicas extraen con mayor claridad las características de las muestras y de esta forma facilitan su clasificación y análisis. A continuación se mencionan algunos estudios realizados para optimizar el análisis de muestras en este dominio.

#### **Análisis de semen de cerdo**

**Enrique Alegre, Lidia Sánchez, Rocía Alaiz, J. Carlos Domínguez [ALE04]**

Este trabajo consistió en la realización de un sistema automático en Matlab para determinar la calidad del semen de cerdo utilizando una Red Neuronal Perceptron multicapa con funciones de activación sigmoidales. El proceso se divide en varias etapas:

- Preprocesamiento de las imágenes, en el cual éstas son transformadas a imágenes en escala de grises.

- Segmentación, basado en regiones usando un proceso de Umbralización Otsu, aislando las cabezas de los espermatozoides y eliminando las colas.
- Limpieza de la imagen, eliminando los objetos cerca de los bordes y aquellos cuyo tamaño sea menor al 45% de la media total de las áreas de los objetos de la imagen.
- Caracterización de las cabezas de los espermatozoides usando descriptores de momentos, Hu y Zernike. Momentos de Hu es la geometría de una región plana basada en el tamaño, posición, orientación y la forma del objeto. Zernike son momentos ortogonales, es decir, se usan bases ortogonales para la extracción de información.

### **Clasificación intracelular del espermatozoide**

Lidia Sánchez González y Enrique Alegre Gutiérrez [SAN05]

Su trabajo se basa en la utilización de descriptores de textura conocidos en la clasificación de células espermáticas en función de la distribución intracelular que presentan. Se utilizan variables estadísticas de segundo orden derivados de la matriz de concurrencia y se comparan los resultados con los obtenidos utilizando características geométricas. Para la clasificación se utiliza el algoritmo de los K-Vecinos más cercanos. Los resultados indican que los descriptores derivados de la matriz de concurrencia clasifican este tipo de imágenes con un error menor que los derivados de las características geométricas.

## **Segmentación automática de espermatozoides**

**Carlos Platero [PLA04]**

El autor elaboró un prototipo de segmentación automática de imágenes de espermatozoides mediante técnicas de difusión isotrópicas y anisotrópicas. Este proceso ejecuta diversas etapas, las cuales abarcan desde la toma de imágenes hasta que la segmentación es realizada:

- Primera etapa, busca las regiones de interés mediante la separación de los espermatozoides respecto del fondo. La transformación del color a una única dimensión que incremente el máximo contraste, entre los objetos de interés y el fondo es aplicado. El análisis del histograma, el uso de un umbral global y la aplicación de técnicas de morfología producirán las regiones de interés. Éstas serán zonas con presencias de halo con uno o varios núcleos.
- Segunda etapa de procesamiento de difusión isotrópica y anisotrópica de cada región de interés resaltando los bordes de la cabeza del espermatozoide. Tras ella, se procede a un análisis local y a una segmentación inicial del núcleo o núcleos existentes en cada región de interés.

### **3.3 Propuesta de investigación**

Con el fin de apoyar a los profesionales que realizan el análisis de morfología espermática, se propone el desarrollo de un sistema automatizado que sea capaz de analizar un conjunto de imágenes digitales que representan diferentes sectores de una muestra de semen, permitiendo generar un resultado general para la muestra. Para realizar esta tarea será necesario construir un modelo de clasificación que tome como entrada las imágenes

preprocesadas, para lo cual se plantea utilizar técnicas de minería de datos y realizar una comparación entre diferentes métodos de clasificación. La aplicación estará dirigida a la evaluación de las cabezas de espermatozoides humanos.

### **3.4 Objetivos**

#### **3.4.1 Objetivo general**

Aplicar la minería de datos para desarrollar un sistema automatizado que realice los procesos de análisis y clasificación de la morfología espermática, a partir de imágenes digitales de muestras de líquido seminal.

#### **3.4.2 Objetivos específicos**

1. Determinar el modelo de clasificación de la morfología de cabeza de espermatozoide humano aplicando técnicas de minería de datos.
2. Desarrollar el sistema automatizado que permita al experto la carga de imágenes y generación de resultados.
3. Verificar el sistema automatizado desarrollado conjuntamente con los expertos.

### **3.5 Estimación del modelo de clasificación**

Un modelo de clasificación permite la asignación de una etiqueta de clase a un registro. Al momento de realizar el análisis morfológico espermático es necesario utilizar un modelo de clasificación confiable, que permita clasificar correctamente a los espermatozoides de la muestra. Debido a que existen diversos métodos de clasificación es necesario determinar cuál es el mejor modelo y realizar diversas pruebas utilizando como base un conjunto de

datos sobre muestras de liquido seminal proporcionados por el experto. En este proceso se hará uso de los algoritmos de minería de datos mostrados en el Capítulo 2 y de la herramienta WEKA, la cual es un software para aprendizaje automático y minería de datos escrito en Java [KIR08].

### 3.5.1 Descripción de la Base de Datos

Para la construcción de un modelo de clasificación fue necesario obtener un conjunto de datos que provean información relacionada al dominio del problema, para esto se contó con el apoyo de un experto, el cual evaluó un conjunto de imágenes digitales que representan diferentes sectores de una muestra de semen mediante el uso de una herramienta previamente desarrollada [FER08].

El experto analizó 220 imágenes, las cuales pertenecen a 12 muestras de semen, en éstas se encontraron 2322 espermatozoides, 144 normales y 2178 anormales. Con los datos obtenidos se creó una base de datos, la cual consta de un conjunto de atributos que caracterizan los espermatozoides y se mencionan a continuación:

1. **Área Cabeza:** Número de píxeles contenidos en la región de la cabeza.
2. **Área Acrosoma:** Número de píxeles contenidos en la región acrosómica.
3. **Área Post-Acrosoma:** Número de píxeles contenidos en la región post-acrosómica.
4. **Eje Horizontal:** Longitud en píxeles del eje horizontal de la cabeza.

5. **Eje Vertical:** Longitud en píxeles del eje vertical de la cabeza.
6. **Elipticidad:** División entre el eje horizontal y eje vertical de la cabeza.
7. **Diferencia entre Cabeza y Elipse Ideal:** Diferencia en píxeles entre el contorno real y el contorno de la elipse ideal calculada a partir de los ejes de la cabeza.
8. **Perímetro Ideal:** Perímetro elíptico definido por:  
$$p = 2 \pi \sqrt{(a^2 + b^2)/2}.$$
9. **Perímetro Real:** Número de píxeles que conforman el borde de la cabeza.
10. **Diferencia Perímetro Ideal-Real:** Diferencia entre perímetro elíptico ideal y perímetro real.
11. **Presencia Pieza Intermedia:** Indica si la pieza intermedia está presente o no.
12. **Grosor Máximo Pieza Intermedia:** Ancho máximo en píxeles de la pieza intermedia.
13. **Ángulo Inserción Pieza Intermedia:** Ángulo en grados entre eje horizontal de la cabeza y eje Pieza Intermedia-Cabeza.
14. **Clasificación:** Es la clasificación asignada al espermatozoide después del análisis.

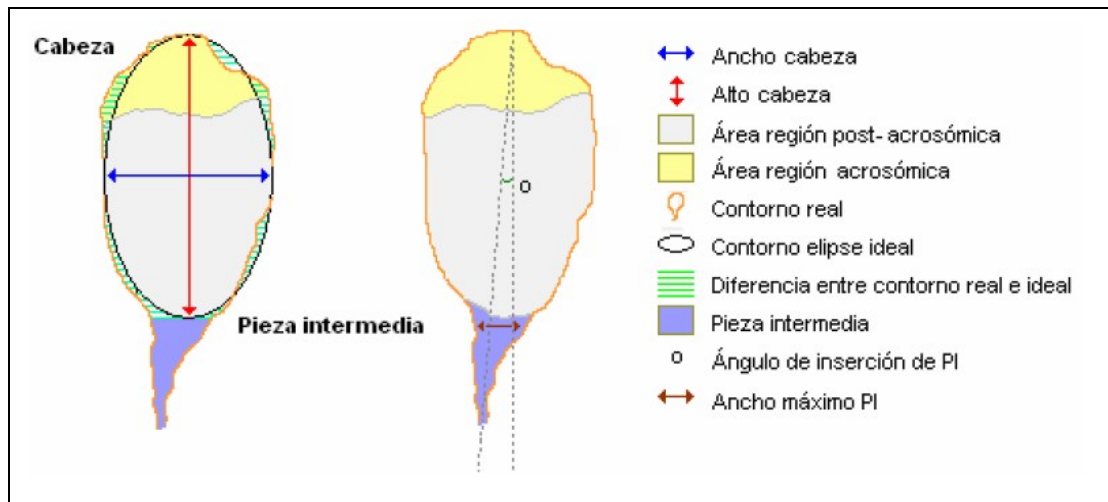


Figura 8: Áreas del espermatozoide.

### 3.5.2 Limpieza y preprocesamiento de los datos

Antes de pasar a la fase de generación del modelo de clasificación, es necesario realizar una limpieza y preprocesamiento de la base de datos. La primera tiene por objetivo detectar valores fuera de rango, valores inconsistentes, valores ausentes y posibles errores, mientras la segunda busca mejorar la representación de los datos para el algoritmo de minería e incluye la creación de nuevos atributos y la selección de variables. Por otra parte es necesario aumentar el número de datos de espermatozoides normales, ya que existe un desbalance de los datos, es decir, existen más datos de espermatozoides cuya clasificación es anormal que normal lo cual dificulta la generación un buen modelo de clasificación.

**Limpieza de los datos:** La imagen segmentada de cada uno de los registros de la base de datos fue inspeccionada con el fin de determinar si la segmentación había sido realizada correctamente y si las características habían sido generadas de acuerdo a lo esperado. Se pudieron detectar

errores en la segmentación para algunos registros, así como errores en el cálculo de algunas características, principalmente en la determinación de los ejes, además, los histogramas de cada una de las variables dieron información de algunos valores extremos; todos estos datos fueron eliminados de la base de datos.

Una vez realizada la limpieza de la base de datos, se pasó a la fase de preprocesamiento, la cual consistió en la eliminación, creación y selección de variables, como se explica a continuación:

**Eliminación de atributos:** Debido a que sólo se realizará la evaluación de la cabeza del espermatozoide los atributos *presencia pieza intermedia*, *grosor máximo de la pieza intermedia* y *el ángulo de inserción pieza intermedia* no fueron usados para la estimación del modelo, ya que éstas se encuentran relacionadas con la pieza intermedia del espermatozoide.

**Creación de nuevas atributos:** Se generó un nuevo atributo que representa el cociente entre los atributos Área Acrosoma y Área Cabeza, ya que según la OMS [OMS01], el acrosoma de un espermatozoide normal debería ocupar entre 40% y 70% de área de la cabeza.

**Selección de atributos:** Para realizar la selección de atributos se hizo uso de algoritmos de selección de atributos ofrecidos por la herramienta WEKA. A continuación se describen los algoritmos usados y las variables seleccionadas por estos [HAL09]:

1. **CfsSubsetEval:** Éste algoritmo evalúa subconjuntos de atributos considerando la habilidad predictiva individual de cada variable. Las variables seleccionadas fueron las siguientes: *Área Post*

*Acrosoma, Cociente Área Acrosoma/Área Cabeza, Eje Horizontal, Eje Vertical, Elipticidad, Perímetro Ideal, Perímetro Real.*

2. **ChiSquareAttributeEval:** Evalúa el valor de cada atributo calculado el valor estadístico Chi-squared. Las variables seleccionadas fueron las siguientes: *Área Post Acrosoma, Eje Horizontal, Eje Vertical, Elipticidad, Perímetro Real.*
  
4. **FilteredAttributeEval:** Con este algoritmo las variables pasan por un filtro el cual evalúa la ganancia de información. Las variables seleccionadas fueron las siguientes: *Área Post Acrosoma, Eje Horizontal, Eje Vertical, Elipticidad, Perímetro Ideal, Perímetro Real.*
  
5. **InfoGain:** Evalúa el valor de cada variable midiendo la ganancia de información con respecto a la clase. Las variables seleccionadas fueron las siguientes: *Área Post Acrosoma, Eje Horizontal, Elipticidad, Perímetro Real.*
  
6. **OneRAttributeVal:** Algoritmo que busca variables con un error mínimo para predicciones y discretización de atributos. Las variables seleccionadas fueron las siguientes: *Cociente Área Acrosoma/Área Cabeza, Eje Horizontal, Elipticidad, Perímetro Ideal, Perímetro Real.*

### 3.5.3 El problema de las clases desbalanceadas

Uno de los factores que suelen enfrentar los algoritmos de minería de datos es el problema de las clases desbalanceadas; esto ocurre cuando se tienen

muchos ejemplos de una clase, pero muy pocos de otra, lo que dificulta generar un buen modelo para clasificar las clases [MOR07].

En el repertorio de datos después del preprocesamiento se obtuvieron 1001 espermatozoides, de los cuales un 87.71% de espermatozoides tienen una clasificación anormal y un 12.29% de espermatozoides con la clasificación normal, debido a esto existe un desbalanceo de la base de datos, el cual ocasiona que al momento de generar el modelo de clasificación exista un gran margen de error en la clasificación de espermatozoides normales. Para solventar esto existen diferentes algoritmos que son capaces de generar datos de la clase minoritaria a partir de los datos existentes, uno de estos algoritmos es el denominado SMOTE [MOR07].

SMOTE crea nuevos datos utilizando elementos existentes de la clase minoritaria. Para lograrlo se debe seguir el siguiente conjunto de pasos:

- Se define porcentaje de ejemplos a generar.
- Para cada ejemplo de la clase minoritaria calcular sus K-Vecinos más cercanos.
- Para cada atributo del ejemplo, calcula la diferencia entre el vector de atributos muestra y el vecino elegido.
- Multiplica esta diferencia por un número aleatorio entre 0 y 1.
- Generar un nuevo elemento cuyos valores son la suma de este último valor al valor original de la muestra.

SMOTE es la herramienta más utilizada para realizar el sobre muestreo pero presenta los siguientes inconvenientes:

- Puede generar muchos ejemplos artificiales cuyas semillas son ejemplos con ruido.

- Al generar un nuevo ejemplo, interpola entre dos ejemplos de la clase minoritaria; sin embargo, pueden existir muchos ejemplos cercanos o inclusive entre ellos de la clase mayoritaria, generando modelos incorrectos.
- Sólo funciona con variables continuas.
- No tiene una forma clara de decidir cuantos ejemplos generar.

Con SMOTE se generaron 500 espermatozoides normales, aumentando la cantidad de 123 a 623.

### **3.5.4 Experimentos y resultados**

Para poder determinar el algoritmo a utilizar en el sistema automatizado fue necesario realizar un conjunto de pruebas, estas evaluaron el desempeño de los algoritmos utilizando diferentes conjuntos de variables. Para este proceso se hizo uso de la herramienta WEKA, la cual ofreció información asociada a los diferentes algoritmos utilizados. Los algoritmos utilizados fueron:

1. K-Vecinos más cercanos (Knn), con  $k=3$  y  $k=5$ .
2. Redes Neuronales (Perceptron Multicapa).
3. Árboles de Decisión (C4.5).
4. Reglas de Clasificación (Ripper).

Para cada experimento se cuenta con un subconjunto de atributos (seleccionados por alguno de los algoritmos de la herramienta WEKA previamente mencionados), con el cual son generados los modelos para cada algoritmo a utilizar. Se generaran 2 modelos por algoritmo, uno tomado los datos previos al balanceo y otro con los datos generados con el balanceo utilizando SMOTE, de esta manera se podrá verificar si hubo una mejora al balancear los datos.

Para cada modelo se calcularán los valores promedios del porcentaje de aciertos, porcentaje de fallos y porcentaje de error utilizando validación cruzada con 10 particiones, y los resultados serán mostrados en tablas; un acierto es un resultado igual al obtenido por el experto. A continuación se presentan las pruebas realizadas y sus resultados.

**Experimento 1:** Para este experimento se hace uso de todos los atributos.

Algoritmo	Resultado para el modelo obtenido con los datos originales			Resultado para el modelo obtenido con los datos balanceados			Resultado del modelo obtenido con los datos balanceados sobre los datos originales		
	AN	AA	Error	AN	AA	Error	AN	AA	Error
<b>Knn (K=3)</b>	33.33	93.05	14.28	94.06	81.89	13.05	90.24	81.89	17.08
<b>Knn (K=5)</b>	30.08	94.64	13.28	94.86	78.92	14.45	87.80	78.92	13.24
<b>Perceptron Multicapa</b>	40.65	94.30	12.28	91.65	82.11	13.92	87.80	82.11	17.18
<b>C4.5</b>	28.45	96.01	12.78	91.49	83.25	13.32	91.05	83.25	15.78
<b>Ripper</b>	14.63	95.33	14.58	85.71	81.54	16.72	85.36	81.54	17.98

Tabla 1 - AN: Aciertos Normales; AA: Aciertos Anormales

**Experimento 2:** Para este experimento se hace uso del algoritmo de selección de atributos: CfsSubsetEval. Los atributos seleccionados: Área Post Acrosoma, Cociente Área Acrosoma y Área Cabeza, Eje Horizontal, Eje Vertical, Elipticidad, Perímetro Ideal y Perímetro Real.

	Resultado para el	Resultado para el	Resultado del modelo
--	-------------------	-------------------	----------------------

Algoritmo	modelo obtenido con los datos originales			modelo obtenido con los datos balanceados			obtenido con los datos balanceados sobre los datos originales		
	AN	AA	Error	AN	AA	Error	AN	AA	Error
<b>Knn (K=3)</b>	39.04	90.27	14.48	91.81	82.91	13.39	94.30	82.91	15.68
<b>Knn (K=5)</b>	35.77	93.73	13.38	93.90	82	13.05	88.61	82	17.18
<b>Perceptron Multicapa</b>	25.20	95.67	12.98	91.49	78.01	16.38	87.80	78.01	20.77
<b>C4.5</b>	0.05	98.74	12.68	92.61	81.43	13.92	85.07	81.43	17.68
<b>Ripper</b>	32.52	95.44	12.28	85.07	83.25	15.98	91.05	83.25	15.78

Tabla 2 - AN: Aciertos Normales; AA: Aciertos Anormales

**Experimento 3:** Para este experimento se emplea del algoritmo de selección de atributos: ChiSquareAttributeEval. Los atributos seleccionados son Área Post Acrosoma, Eje Horizontal, Eje Vertical, Elipticidad y Perímetro Real.

Algoritmo	Resultado para el modelo obtenido con los datos originales			Resultado para el modelo obtenido con los datos balanceados			Resultado del modelo obtenido con los datos balanceados sobre los datos originales		
	AN	AA	Error	AN	AA	Error	AN	AA	Error
<b>Knn (K=3)</b>	29.26	92.82	14.98	91.88	82.57	14.39	80.48	82.57	17.68
<b>Knn (K=5)</b>	29.26	92.82	14.98	91.65	80.41	14.92	84.55	80.41	19.08
<b>Perceptron Multicapa</b>	16.26	97.38	12.58	90.20	77.79	17.05	78.04	77.79	17.05
<b>C4.5</b>	9.75	97.26	13.48	89.24	84.16	13.72	86.17	84.16	15.28
<b>Ripper</b>	90.24	81.89	13.18	79.61	84.28	17.65	82.92	84.28	15.88

Tabla 3 - AN: Aciertos Normales; AA: Aciertos Anormales

**Experimento 4:** Para este experimento se hace uso del algoritmo de selección de atributos: FilteredAttributeEval. Los atributos

seleccionados son Área Post Acrosoma, Eje Horizontal, Eje Vertical, Elipticidad, Perímetro Ideal y Perímetro Real.

Algoritmo	Resultado para el modelo obtenido con los datos originales			Resultado para el modelo obtenido con los datos balanceados			Resultado del modelo obtenido con los datos balanceados sobre los datos originales		
	AN	AA	Error	AN	AA	Error	AN	AA	Error
<b>Knn (K=3)</b>	30.89	92.59	14.98	89.72	83.14	14.12	80.48	82.57	15.48
<b>Knn (K=5)</b>	48.19	94.76	12.88	91.49	80.86	14.72	88.61	80.86	18.18
<b>Perceptron Multicapa</b>	15.47	81.89	13.18	90.20	78.47	16.65	91.05	78.47	19.98
<b>C4.5</b>	9.75	97.38	19.98	87.80	84.96	13.85	92.68	84.96	14.08
<b>Ripper</b>	23.57	96.92	12.08	83.94	85.19	15.32	82.92	85.19	15.08

Tabla 4 - AN: Aciertos Normales; AA: Aciertos Anormales

**Experimento 5:** Para este experimento se hace uso del algoritmo de selección de atributos: InfoGain. Los atributos seleccionados son Área Post Acrosoma, Eje Horizontal, Elipticidad y Perímetro Real.

Algoritmo	Resultado para el modelo obtenido con los datos originales			Resultado para el modelo obtenido con los datos balanceados			Resultado del modelo obtenido con los datos balanceados sobre los datos originales		
	AN	AA	Error	AN	AA	Error	AN	AA	Error
<b>Knn (K=3)</b>	30.89	92.71	14.88	91.97	81.20	14.32	78.86	81.20	19.08
<b>Knn (K=5)</b>	30.89	94.19	13.58	92.93	79.95	14.65	83.73	79.95	19.58
<b>Perceptron Multicapa</b>	21.95	96.92	12.28	89.72	77.10	26.47	89.43	77.10	12.28
<b>C4.5</b>	7.31	98.40	12.78	89.40	82	14.92	90.24	82	16.98
<b>Ripper</b>	18.69	96.12	13.38	86.35	82.68	15.78	91.05	82.68	15.78

Tabla 5 - AN: Aciertos Normales; AA: Aciertos Anormales

**Experimento 6:** Para este experimento se hace uso del algoritmo de selección de atributos: OneRAttributeVal. Los atributos seleccionados

son Cociente Área Acrosoma y Área Cabeza, Eje Horizontal, Elipticidad, Perímetro Ideal y Perímetro Real

Algoritmo	Resultado para el modelo obtenido con los datos originales			Resultado para el modelo obtenido con los datos balanceados			Resultado del modelo obtenido con los datos balanceados sobre los datos originales		
	AN	AA	Error	AN	AA	Error	AN	AA	Error
<b>Knn (K=3)</b>	31.70	92.82	14.68	91.34	81.43	14.12	82.92	81.43	18.38
<b>Knn (K=5)</b>	25.20	95.10	13.48	94.22	80.06	19.08	86.99	80.06	19.08
<b>Perceptron Multicapa</b>	25.20	96.12	12.58	89.88	77.56	17.32	87.80	77.56	21.17
<b>C4.5</b>	1.62	99.31	12.68	90.52	83.25	15.88	90.24	83.25	15.88
<b>Ripper</b>	29.26	95.33	12.78	89.40	81.89	14.99	88.61	81.89	17.28

Tabla 6 - AN: Aciertos Normales; AA: Aciertos Anormales

En los experimentos se puede observar que los resultados obtenidos de los modelos generados con los datos no balanceados, no presentan porcentajes de aciertos altos al clasificar los espermatozoides normales, mientras que para los espermatozoides anormales sí se obtuvo un alto porcentaje de acierto; a su vez, para los modelos obtenidos con los datos balanceados el porcentaje de aciertos de espermatozoides normales es más alto y hubo un leve descenso en los aciertos de los espermatozoides anormales.

Después de evaluar los diferentes resultados obtenidos, el algoritmo seleccionado para desarrollar el modelo de clasificación fue Knn, con k igual a 3, debido a que el Experimento 2 obtuvo buenos resultados, con una tasa de error de 15.68%; los atributos usados en este experimente fueron: Área Post Acrosoma, Cociente Área Acrosoma y Área Cabeza, Eje Horizontal, Eje Vertical, Elipticidad, Perímetro Ideal y Perímetro Real.

### 3.6 Desarrollo de la aplicación

Esta sección describe las etapas de desarrollo del sistema automatizado: análisis, diseño, aspectos relevantes del desarrollo y pruebas.

### **3.6.1 Análisis de requerimientos**

Se requiere desarrollar un sistema que automatice el proceso de análisis morfológico. El sistema debe ser capaz de cargar, segmentar y procesar imágenes de muestra de líquido seminal, para luego clasificarlas con el mejor modelo obtenido con las técnicas de minería de datos. Para esto se deben plantear un conjunto de requerimientos que abarquen el dominio de la situación, estos son:

#### **Requerimientos funcionales**

- Permitir la carga de imágenes de las muestras de semen.
- Permitir al usuario establecer parámetros, tales como la cantidad de espermatozoides a contar.
- Realizar la clasificación de la morfología espermática de las muestras cargadas por el usuario.
- Generar y mostrar resultados de la clasificación obtenida.

#### **Requerimientos no funcionales**

- Utilizar el lenguaje de programación JAVA.

### **3.6.2 Modelo de Casos de uso**

A continuación se presenta un diagrama de casos de uso que muestra una representación gráfica del entorno del sistema (actores) y sus funcionalidades principales (Figura 8).

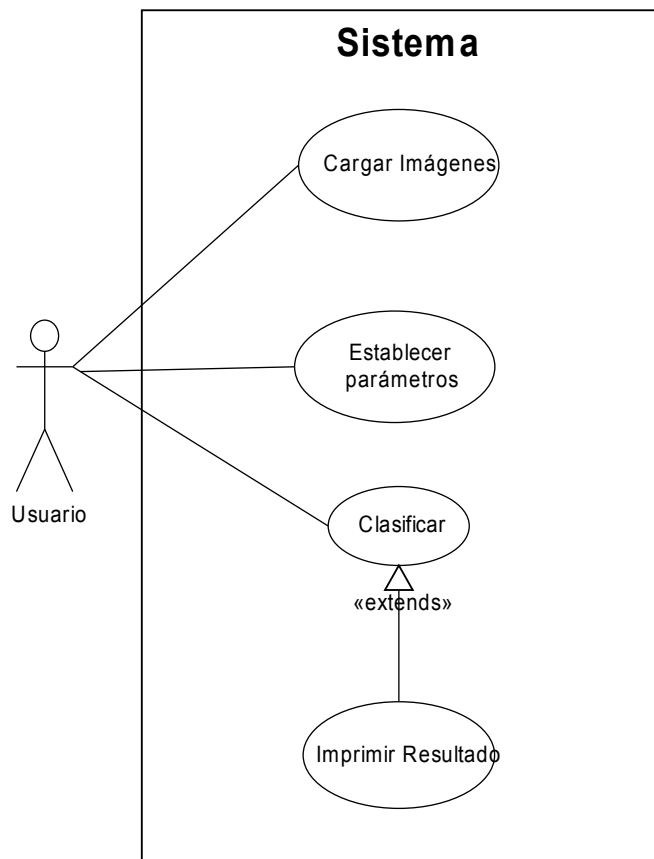


Figura 9: Diagrama de Casos de Usos

### Descripción

<b>Nombre</b>	Cargar Imágenes
<b>Actor</b>	Usuario
<b>Descripción</b>	Permite al usuario seleccionar una imagen o una carpeta de imágenes para cargarla en el sistema.
<b>Flujo de Eventos</b>	<ol style="list-style-type: none"> <li>1.- El actor selecciona la opción Abrir en el Menú Archivo en la barra de tareas.</li> <li>2.- El actor selecciona la imagen o carpeta de imágenes.</li> <li>3.- El Sistema carga la(s) imagen(es) seleccionada(s).</li> <li>4.- El Sistema muestra la(s) imagen(es) cargadas.</li> </ol>
<b>Pre-condición</b>	---
<b>Post-condición</b>	Las imágenes fueron seleccionadas.

<b>Nombre</b>	Establecer Parámetros
<b>Actor</b>	Usuario
<b>Descripción</b>	Permite al usuario configurar la cantidad de espermatozoides a contar y el porcentaje mínimo a cumplir para que el resultado del análisis sea Normal.
<b>Flujo de Eventos</b>	<ol style="list-style-type: none"> <li>1.- El actor selecciona la opción Configurar en el Menú opciones o en la barra de tareas.</li> <li>2.- Se muestra una ventana de configuración.</li> <li>3.- El usuario especifica la cantidad de espermatozoides a contar.</li> <li>4.- El usuario especifica el porcentaje mínimo a cumplir para que el resultado sea normal.</li> </ol>
<b>Pre-condición</b>	---
<b>Post-condición</b>	Se establece la cantidad de espermatozoides a contar y el porcentaje mínimo a cumplir para que el resultado del análisis sea normal.

<b>Nombre</b>	Clasificar
---------------	------------

<b>Actor</b>	Usuario
<b>Descripción</b>	Funcionalidad principal del sistema, a través de la cual se lleva a cabo el proceso de análisis de una imagen o conjunto de imágenes.
<b>Flujo de Eventos</b>	<p>1.- El usuario selecciona la opción Ejecutar en el Menú de opciones o en la barra de tareas.</p> <p>2.- El Sistema lleva a cabo el proceso de análisis con las imágenes cargadas previamente por el usuario.</p> <p>3.- El Sistema muestra los resultados.</p>
<b>Pre-condición</b>	El usuario debió haber ejecutado el caso de uso <b>Cargar Imágenes</b> .
<b>Post-condición</b>	El usuario obtiene el resultado del análisis morfológico espermático a partir de las imágenes digitales dadas.

<b>Nombre</b>	Imprimir Resultado
<b>Actor</b>	Usuario
<b>Descripción</b>	Permite al usuario Imprimir los resultados del procesamiento de archivos.
<b>Flujo de Eventos</b>	<p>1.- El actor selecciona la opción Imprimir que se encuentra en el menú Archivo del Sistema.</p> <p>2.- El Sistema Imprimir los resultados obtenidos.</p>
<b>Pre-condición</b>	El sistema debe haber ejecutado el caso de uso <b>Clasificar</b> .
<b>Post-condición</b>	El usuario obtiene la impresión de los resultados obtenidos.

### 3.6.3 Arquitectura de la aplicación

Para poder automatizar el análisis de la morfología espermática, es necesario definir un conjunto de módulos, estos abarcan desde la comunicación con el usuario hasta el resultado final, a continuación se presenta una breve descripción de los diferentes módulos de la aplicación (Figura 9):

1. **Interfaz Usuario:** Es el conjunto de componentes que representan los elementos visuales que permiten al usuario acceder a las funcionalidades del sistema y realizar la selección de imágenes de la muestra que se desea analizar.
2. **Segmentador:** Módulo encargado de realizar las tareas de segmentación de las muestras provistas por el usuario, como salida se obtienen un conjunto de imágenes segmentadas que representan los diferentes espermatozoides de la muestra seleccionada. Para lograr esto se carga una imagen o un conjunto de imágenes al sistema, estas imágenes son preprocesadas y luego divididas utilizando un conjunto de criterios que permiten separar los sectores útiles para el dominio de la aplicación. Este módulo está relacionado con el proyecto *Herramienta para la extracción automática de características morfológicas de espermatozoides humanos a partir de imágenes* [FER08].
3. **Extractor de Características:** Este módulo se encarga de obtener las características de las diferentes imágenes de espermatozoides obtenidos del **Segmentador**. Se hace uso de un conjunto de filtro y algoritmos para evaluar de diferentes maneras las imágenes. Este módulo está relacionado con el proyecto *Herramienta para la extracción automática de características morfológicas de espermatozoides humanos a partir de imágenes* [FER08].
4. **Clasificador:** Componente encargado de clasificar cada espermatozoide obtenido de la muestra, utilizando el modelo seleccionado en la Sección 3.5.4.

5. **Analizador:** Genera un resultado general de las muestras dependiendo del conteo obtenido en la clasificación. El resultado es dado tomando en cuenta las reglas de la OMS.

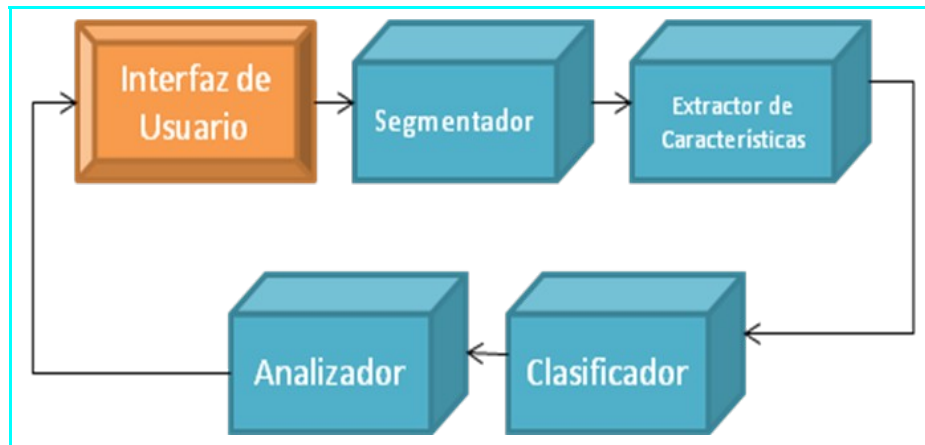


Figura 10: Módulos de la aplicación

### 3.6.4 Plataforma tecnológica

Para la construcción de la aplicación, se decidió hacer uso del lenguaje de programación **JAVA**, debido a que es un lenguaje de programación orientado permite al sistema ser multiplataforma, modular y escalable.

La aplicación fue desarrollada usando **Eclipse**, entorno de desarrollo integrado de código abierto multiplataforma para desarrollar proyectos, bajo el sistema operativo **Windows** [ECL09].

### 3.6.5 Interfaz de la aplicación

Para el sistema automatizado se diseñó una interfaz que permitiera al usuario, de manera fácil e intuitiva, realizar la carga de imágenes, iniciar el proceso de análisis morfológico, ver los resultados y configurar el sistema. En la Figura 10 se presenta un modelo de la interfaz.

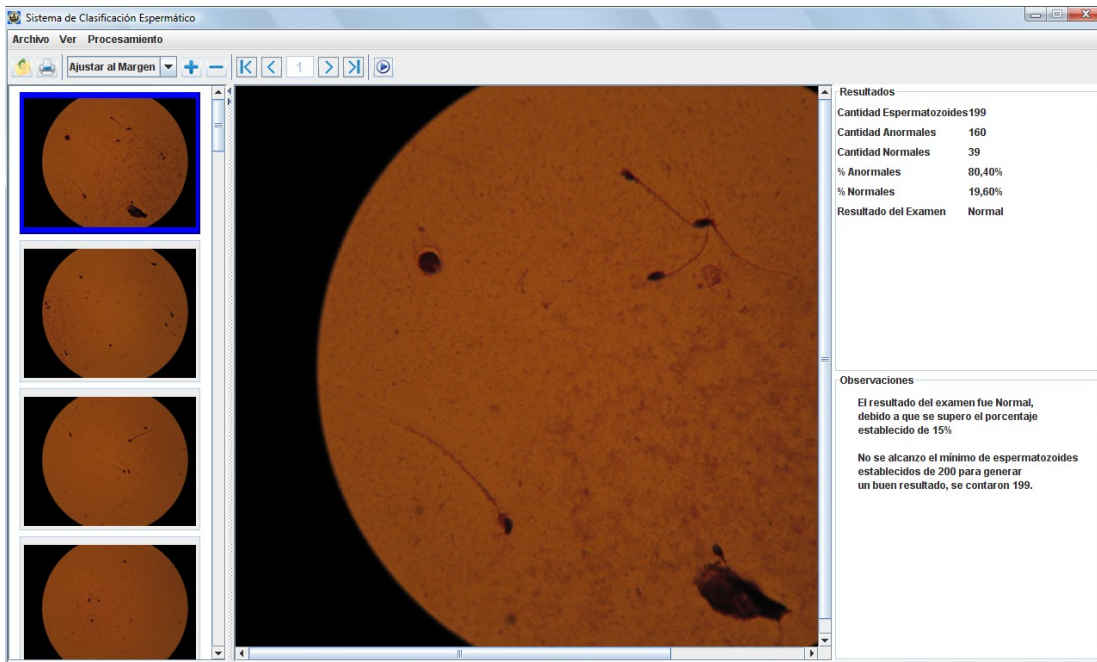


Figura 11: Interfaz de la Aplicación

- La parte superior representa el menú de la aplicación, en éste se encuentran las diferentes opciones que el usuario puede ejecutar, entre ellas el inicio del procesamiento de las muestras.



Figura 12: Menú de la aplicación

- El área de la izquierda representa el listado de imágenes seleccionadas para analizar.

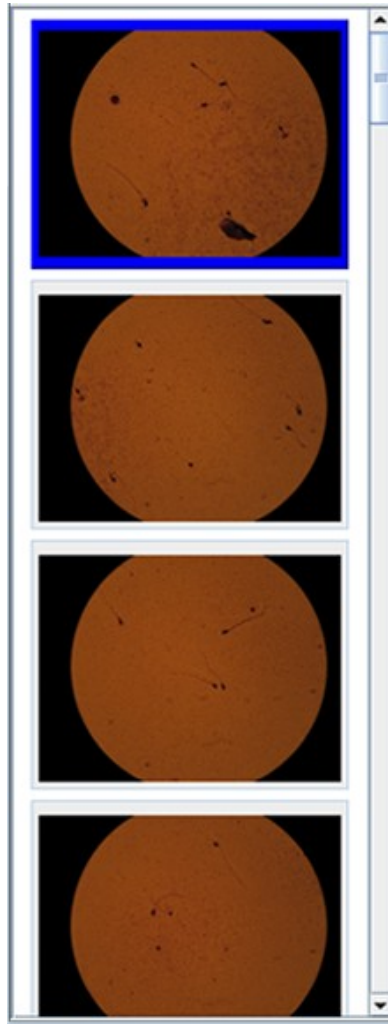
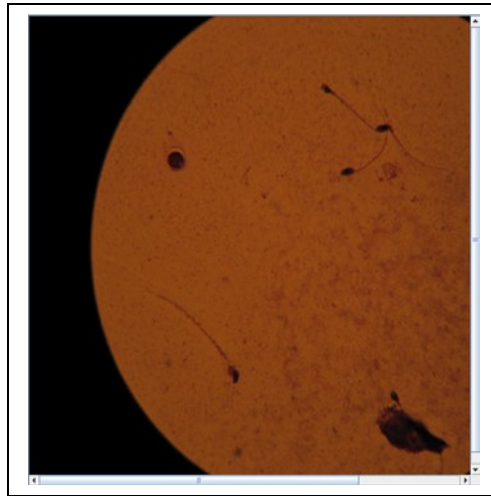


Figura 13: Listado de imágenes seleccionadas

- El área central muestra la imagen seleccionada del listado de imágenes



. Figura 14: Listado de imágenes seleccionadas

- El área derecha muestra los resultados obtenidos del análisis.

Resultados	
Cantidad Espermatozoides 199	
Cantidad Anormales	160
Cantidad Normales	39
% Anormales	80,40%
% Normales	19,60%
Resultado del Examen	Normal

Observaciones
El resultado del examen fue Normal, debido a que se supero el porcentaje establecido de 15%
No se alcanzo el mínimo de espermatozoides establecidos de 200 para generar un buen resultado, se contaron 199.

Figura 15: Listado de imágenes seleccionadas

### 3.6.6 Verificación de la aplicación

Una vez culminada la aplicación se realizaron pruebas para verificar que los resultados obtenidos sean correctos; para esto se tomaron las imágenes analizadas por el experto y se procesaron con la aplicación, con el fin de verificar que los resultados coincidan con los resultados del experto.

El experto analizó 220 imágenes, las cuales pertenecen a 12 muestras de semen, en éstas encontró 2322 espermatozoides, de los cuales 144 clasificó como normales y 2178 anormales.

El Sistema Automatizado analizó las mismas imágenes que usó el experto y obtuvo un total de 2378 espermatozoides, 520 normales y 1918 anormales. El Sistema encontró más espermatozoides que el experto, esto posiblemente se deba a posibles problemas durante el proceso de segmentación y a que el experto durante el análisis descartó algunos espermatozoides por no cumplir con los requisitos mínimos para ser evaluado.

A continuación se muestra una tabla con los resultados obtenidos, mostrando la cantidad de espermatozoides anormales y normales contados y el resultado final del examen, tomando en cuenta como porcentaje de Terazoospermia 15% para cada una de las 12 muestras.

	<b>Resultados del experto</b>	<b>Resultados de la aplicación</b>
<b>Muestra 1</b>	17 Normales	39 Normales

	148 Anormales Resultado: Terazoospermia	160 Anormales Resultado: Normal
<b>Muestra 2</b>	44 Normales 162 Anormales Resultado: Terazoospermia	83 Normales 117 Anormales Resultado: Normal
<b>Muestra 3</b>	15 Normales 164 Anormales Resultado: Terazoospermia	43 Normales 154 Anormales Resultado: Normal
<b>Muestra 4</b>	3 Normales 225 Anormales Resultado: Terazoospermia	14 Normales 186 Anormales Resultado: Terazoospermia
<b>Muestra 5</b>	5 Normales 199 Anormales Resultado: Terazoospermia	21 Normales 179 Anormales Resultado: Terazoospermia
<b>Muestra 6</b>	16 Normales 192 Anormales Resultado: Terazoospermia	85 Normales 115 Anormales Resultado: Normal
<b>Muestra 7</b>	0 Normales 208 Anormales Resultado: Terazoospermia	14 Normales 186 Anormales Resultado: Terazoospermia
<b>Muestra 8</b>	2 Normales 135 Anormales Resultado: Terazoospermia	16 Normales 184 Anormales Resultado: Terazoospermia
<b>Muestra 9</b>	23 Normales 206 Anormales Resultado: Terazoospermia	44 Normales 156 Anormales Resultado: Normal
<b>Muestra 10</b>	15 Normales 197 Anormales Resultado: Terazoospermia	83 Normales 117 Anormales Resultado: Normal
<b>Muestra 11</b>	4 Normales 173 Anormales Resultado: Terazoospermia	58 Normales 131 Anormales Resultado: Normal
<b>Muestra 12</b>	0 Normales 169 Anormales Resultado: Terazoospermia	20 Normales 173 Anormales Resultado: Terazoospermia

Tabla 6: Resultados obtenidos por el experto y la aplicación.

Los resultados muestran que el 41% de los resultados de los exámenes obtenidos por el Sistema coinciden con los resultados del experto; este bajo porcentaje de acierto puede deberse al margen de error de 15% existente en el modelo mostrado en el experimento 2, a errores durante el proceso de segmentación y debido al bajo porcentaje de espermatozoides normales requeridos para determinar si una muestra presenta Terazoospermia el umbral puede ser fácilmente superado.

## CONCLUSIONES

Vista la necesidad de optimizar el proceso de clasificación morfológica espermática durante el análisis de líquido seminal, se implementó una aplicación que automatiza el proceso de estudio de la morfología del espermatozoide. El Sistema automatizado fue construido con el fin de clasificar muestras de semen a partir de imágenes digitales, para determinar el porcentaje de espermatozoides normales y anormales.

La construcción de la aplicación se llevó a cabo partiendo de los conceptos de minería de datos y segmentación de imágenes digitales y tomando en cuenta el proceso manual del análisis de la morfología espermática. De acuerdo con esto y siguiendo los requerimientos solicitados por el usuario, la aplicación realiza el análisis de las muestras llevando a cabo los siguientes pasos:

- **Carga de imágenes digitales.**
- **Segmentación de imágenes.**
- **Extracción de características.**
- **Clasificación.**

Para la aplicación, se utilizó el algoritmo K-Vecinos más cercanos, con  $k=3$ . Los resultados obtenidos por el modelo generado, previo a la aplicación del algoritmo SMOTE, obtuvieron un 39.04% de aciertos clasificando espermatozoides normales y un 90.27% de acierto en espermatozoides anormales; mientras que el modelo generado luego de aplicar el algoritmo SMOTE obtuvieron un 91.81% de aciertos clasificando espermatozoides normales y un 82.91% de aciertos en espermatozoides anormales. Como se observa el algoritmo SMOTE mejoró considerablemente los resultados para

la clasificación de espermatozoides normales, mientras que para los espermatozoides anormales hubo un descenso en el porcentaje de aciertos, posiblemente debido a que el número de registros generados con SMOTE coinciden con los registros anormales.

El Sistema automatizado, usando el modelo seleccionado, obtuvo un 41% de aciertos en los resultados de los exámenes con respecto a la clasificación obtenida por un experto con el proceso manual. Esta diferencia puede deberse al margen de error existente en el modelo mostrado seleccionado, a errores durante el proceso de segmentación o al bajo porcentaje de espermatozoides normales requeridos para determinar si una muestra presenta Terazoospermia.

En la opinión del autor, el modelo generado obtuvo buenos resultados, a pesar del bajo porcentaje de acierto obtenido por el sistema, con revisar los resultados y evaluar los algoritmos de clasificación se pueden mejorar los resultados con un nuevo modelo.

## **Aportes**

- Este trabajo es una compilación de información asociada al análisis de morfología espermática y minería de datos, la cual puede ser útil en investigaciones donde se requieran conceptos referentes al análisis de líquido seminal y/o empleo de paradigmas de clasificación.

## **Recomendaciones y futuros trabajos**

- Mejorar los procesos de segmentación y extracción de características, con el fin de obtener mejor tiempo de respuesta.
- Incluir información de posible ruido en las imágenes en el proceso de clasificación, con el fin de determinar si los datos son de un espermatozoide o no.

## REFERENCIAS

**[ACO04]** Acosta M. *Minería de Datos y Descubrimiento de Conocimiento (Data Mining and Knowledge Discovery)*, Facultad de Economía, Universidad de La Habana, 2.004. Artículo extraído desde:

<http://www.congreso-info.cu/UserFiles/File/Info/Info2004/Ponencias/056.pdf>

**[ALE04]** Alegre E., Sánchez L., Alaiz R. y Domínguez-Fernández J.C. *Utilización de momentos estadísticos en la clasificación de cabezas de espermatozoides de verraco*. XXV Jornadas de Automática Ciudad Real, 2.004.

**[APT97]** Apte C. y Weiss S. *Data Mining with Decision Trees and Decision Rules*, Future Generation Computer Systems , 1.997. Artículo extraído desde:

[http://reference.kfupm.edu.sa/content/d/a/data\\_mining\\_with\\_decision\\_trees\\_and\\_deci\\_193568.pdf](http://reference.kfupm.edu.sa/content/d/a/data_mining_with_decision_trees_and_deci_193568.pdf)

**[CHA03]** Chanona F., García A., Ruvalcaba C. y Cuneo P. *Necrozoospermia. ¿Mejora con el uso de Metformin?*, Centro de Fertilidad Humana Guadalajara, México, 2.003. Artículo extraído desde:

[http://www.asesa.org/congresos\\_asesa/stdomingo/PDF/MAMP6.pdf](http://www.asesa.org/congresos_asesa/stdomingo/PDF/MAMP6.pdf)

**[FER08]** Ferreira, H. *Herramienta para la extracción automática de características morfológicas de espermatozoides humanos a partir de imágenes*, Trabajo Especial de Grado. Escuela de Computación. Facultad de Ciencias. Universidad Central de Venezuela, 2.008.

**[FLO08]** Flores, L. *Red Neuronal Artificial*. Instituto Tecnológico de Nogales, Área de Ingeniería en Sistemas Computacionales, México, 2.008. Artículo extraído desde:

<http://www.itnogales.edu.mx/formatos/Red%20neuronal%20artificial.pdf>

**[GAR08]** García, M., Quintales L., García F., y Polo M. *Obtención y Validación de modelos de estimación de software mediante técnicas de minería de datos*. Revista Colombiana de Computación, 2.008, Artículo Extraído desde:

[http://www.unab.edu.co/editorialunab/revistas/rcc/pdfs/r31\\_art3\\_c.pdf](http://www.unab.edu.co/editorialunab/revistas/rcc/pdfs/r31_art3_c.pdf)

**[HAL09]** Hall H. *Selección de Atributos*. Pentaho Community, 2.009. Publicación extraída desde:

<http://wiki.pentaho.com/display/DATAMINING/Attribute+Selection>

**[HER05]** Hernansáez J., Botía J. y Gómez-Skarmeta A. *Diseño avanzado de bases de datos*. Dpto. de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, España, 2.005. Artículo extraído desde:

[http://basesdatos.uc3m.es/fileadmin/Docencia/DABDII/Curso\\_2007\\_2008\\_Legans/Tema3\\_01.pdf](http://basesdatos.uc3m.es/fileadmin/Docencia/DABDII/Curso_2007_2008_Legans/Tema3_01.pdf)

**[KIR08]** Kirkby R., Frank E. y Reutemann P. (2008). *WEKA Explorer User Guide*, Universidad de Waikato, Nueva Zelanda, 2.008. Artículo extraído desde:

<http://hivelocity.dl.sourceforge.net/sourceforge/weka/ExplorerGuide-3-5-8.pdf>

**[MOR07]** Morales, E. y González J. *El problema de las clases desbalanceadas*, 2.007, Artículo Extraído desde:

<http://ccc.inaoep.mx/~jagonzalez/ML2/Inbalanced.pdf>

**[OMS01]** – Organización Mundial de la Salud. *Manual de Laboratorio para el examen del semen humano y de la interacción entre el semen y el moco cervical*. Editorial Médica Panamericana S.A. Madrid España. Abril. Cuarta Edición, 2.001.

**[ORA04]** Orallo, J., Quintana, J. y Ferri, C. *Introducción a la Minería de Datos*. Prentice-Hall. 2.004.

**[PLA04]** Platero, Carlos. *Segmentación automática de imágenes de espermatozoides humanos mediante técnicas de difusión isotrópicas y anisotrópicas*. XXV Jornadas de Automática. Ciudad Real, España. 2.004.

**[PLA07]** Pla F., Vázquez F. y Sánchez J. *Una propuesta basada en la estimación de las probabilidades para la edición utilizando el Clasificador kNN*, 2.007. Artículo extraído desde:

[http://marmota.dlsi.uji.es/WebBIB//papers/2007/0\\_Vazquez-MIA-2007.pdf](http://marmota.dlsi.uji.es/WebBIB//papers/2007/0_Vazquez-MIA-2007.pdf)

**[QUI07]** Quintales L., Moreno M., García F. y Polo M. J. *Aplicación de Técnicas de Minería de Datos en la Construcción y Validación de Modelos Predictivos y Asociativos a partir de Especificaciones de Requisitos de Software*, Universidad de Salamanca, 2.007. Artículo extraído desde:

<http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-84/paper4.pdf>

**[SAN05]** Sánchez L. y Alegre E. *Evaluación de Descriptores de Textura en la Clasificación de Células Espermáticas*. XXVI Jornadas de Automática. Alicante-Elche, España, 2.005.

**[VAL02]** Valverde J. *Factores Predictores de Embarazo en Inseminación Intrauterina*, Universidad Complutense de Madrid, España, 2.002 Artículo  
Extraído desde: <http://www.ucm.es/BUCM/tesis/med/ucm-t26149.pdf>

**[VAL06]** Vallejos, S. *Minería de Datos*. Universidad Nacional de Nordeste, Argentina, 2.006, Artículo Extraído desde:  
[http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria\\_Datos\\_Vallejos.pdf](http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf)

**[VAS07]** Vásquez, F. y Vásquez D. *Espermograma y su utilidad clínica*. Salud Uninorte. Barranquilla (Col.) 2007; 23(2): 220-230. Artículo Extraído desde:  
[http://ciruelo.uninorte.edu.co/pdf/salud\\_uninorte/23-2/8\\_Espermograma.pdf](http://ciruelo.uninorte.edu.co/pdf/salud_uninorte/23-2/8_Espermograma.pdf)

**[ECL09]** Eclipse Foundation. *What is Eclipse and the Eclipse Foundation*. Extraído desde:  
<http://www.eclipse.org/org/#about>

**[JOS09]** *Sistemas de Expertos Basados en Reglas*. Dpto. de Matemática de Aplicada. Universidad de Cantabria. Extraído desde:  
<http://personales.unican.es/gutierjm/cursos/expertos/Reglas.pdf>