

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
**UNIVERSIDAD
CATÓLICA**
DEL PERÚ

**ANÁLISIS DE CARACTERÍSTICAS DE FORMA DEL BACILO DE
Koch PARA DETECCIÓN AUTOMÁTICA DE TUBERCULOSIS EN
IMÁGENES DIGITALES**

Tesis para optar el Título de Ingeniero Electrónico, que presenta el bachiller:

JAVIER EDUARDO TICONA HUAROTO

ASESORES: Ph.D Benjamín Castañeda Aphan y Ph.D Roberto Lavarello Montero

Lima, Julio de 2017

RESUMEN

La Tuberculosis es una de las enfermedades más letales a nivel mundial. Los esfuerzos en salud pública están dirigidos a la temprana detección de los casos bacilíferos, ya que son la fuente de infección. En el mundo la detección se realiza mediante baciloscopía, que consiste en la observación de muestras de esputo para identificar y contar bacilos con la ayuda de un microscopio. Sin embargo, el procedimiento es subjetivo y consume excesivo tiempo al personal de salud. El presente estudio tiene como objetivo identificar bacilos en imágenes digitales captadas desde el microscopio.

Dichas imágenes muestran bacilos y otros artefactos con el mismo color. Ambos tipos de estructura se almacenaron y etiquetaron individualmente conformando la base de datos. Se analizó el espectro de magnitudes de los descriptores de Fourier de dichas estructuras, con el fin de seleccionar los necesarios para la óptima caracterización e identificación. Mediante el método sub-óptimo de selección de características hacia atrás (backward feature selection) se determinó los 14 descriptores que mejor discriminan entre las clases. Para comprobar este método se diseñó un programa que procesó las 480 estructuras de la base de datos. Dicho programa obtuvo un porcentaje de acierto de 96.86%, una sensibilidad de 100% y una especificidad de 91.47%

El estudio demuestra que es posible la identificación de bacilos mediante la clasificación de descriptores de Fourier previamente seleccionados. Estos resultados sugieren que las técnicas de procesamiento de imágenes digitales tienen el potencial de agilizar el diagnóstico de Tuberculosis.



PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA

TEMA DE TESIS PARA OPTAR EL TÍTULO DE INGENIERO ELECTRÓNICO

Título : Análisis de Características de Forma del Bacilo de Koch para
Detección Automática de Tuberculosis en Imágenes Digitales

Área : Procesamiento Digital de Imágenes # 605

Asesores : Benjamín Castañeda Aphan
Roberto Lavarello Montero

Alumno : Javier Eduardo Ticona Huaroto

Código : 20010370

Fecha : 17 de Junio de 2008





Descripción y Objetivos

La tuberculosis (TBC) es una enfermedad infecto-contagiosa producida por el Bacilo de Koch. Según la Organización Mundial de la Salud, 8 millones de personas se enferman cada año de TBC, de las cuales morirán de 2 a 3 millones. El 95% de los casos se encuentran en los países pobres como el nuestro. En el Perú, la baciloscopia es la técnica más utilizada para el diagnóstico de TBC. En ella, una muestra de esputo del paciente es procesada utilizando la técnica de tinte Ziehl Neelsen y observada bajo el microscopio para buscar bacilos. Desventajas de esta técnica son la experiencia que se requiere de parte del operador y el tiempo que le toma al operador revisar las muestras.

Una técnica propuesta en la literatura especializada para automatizar la detección de bacilos consiste en utilizar técnicas de procesamiento digital de imágenes basadas en detección de borde y discriminación de color. Sin embargo, estas técnicas obtienen un número alto de falsos positivos. El presente tema de tesis tiene como propósito mejorar la precisión de la detección de bacilos usando procesamiento de imágenes mediante el análisis de las características de forma de los bacilos. Primero, se creará, anotará y validará una base de datos de imágenes microscópicas de esputo bajo la asesoría de técnicos especialistas del Hospital Dos de Mayo. Segundo, se implementarán técnicas matemáticas de descripción de forma a las estructuras obtenidas al aplicar segmentación basada en detección de bordes. Finalmente, se utilizará técnicas de reconocimiento de patrones para diferenciar entre bacilos de Koch y otras estructuras detectadas por los algoritmos de detección basados en detección de bordes.

MÁXIMO 100 PÁGINAS



PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU
SECCION ELECTRICIDAD Y ELECTRONICA


Ing. ANDRÉS FLORES ESPINOZA
Coordinador de la Especialidad de Ingeniería Electrónica



PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA

TEMA DE TESIS PARA OPTAR EL TÍTULO DE INGENIERO ELECTRÓNICO

Título : Análisis de Características de Forma del Bacilo de Koch para
Detección Automática de Tuberculosis en Imágenes Digitales

Índice

Introducción

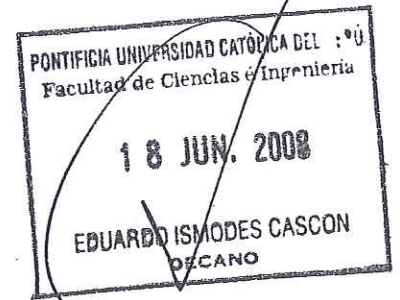
1. Diagnóstico de Tuberculosis: Definición y Problemática
2. Conceptos Generales de Descriptores de Forma
3. Extracción y Análisis de Descriptores de Forma de Bacilos y de Estructuras no Bacilos
4. Aplicación de Reconocimiento de Patrones para la Detección de Bacilos

Conclusiones

Recomendaciones

Bibliografía

Anexos



PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU
SECCION ELECTRICIDAD Y ELECTRONICA

Ing. ANDRES FLORES ESPINOZA
Coordinador de la Especialidad de Ingeniería Electrónica

MÁXIMO 100 PÁGINAS

INDICE DE CONTENIDO

INTRODUCCIÓN	1
1 DIAGNÓSTICO DE TUBERCULOSIS: DEFINICIÓN Y PROBLEMÁTICA	3
1.1 Introducción	3
1.1.1 La tuberculosis	3
1.1.2 La tuberculosis en el mundo	4
1.1.3 La Tuberculosis en el Perú	5
1.1.4 Modo de transmisión	5
1.1.5 Patogenia	6
1.1.6 Síntomas	6
1.2 Diagnóstico	6
1.2.1 Métodos de diagnóstico	7
1.2.1.1 Métodos de diagnóstico de infección	7
1.2.1.2 Métodos de diagnóstico de enfermedad	8
1.3 Definición del problema	12
1.4 Estado del arte	14
1.5 Ventajas de la automatización	19
1.6 Sistema de detección automática de tuberculosis	19
1.7 Objetivos	21
1.7.1 Objetivo principal	22
1.7.2 Objetivos secundarios	22
2 CONCEPTOS GENERALES DE DESCRIPTORES DE FORMA	24
2.1 Modelo teórico	24
2.2 Descriptores de forma	24
2.3 Reconocimiento de patrones	30
2.3.1 Selección de características	33
3 EXTRACCIÓN Y ANÁLISIS DE DESCRIPTORES DE FORMA DE BACILOS Y DE ESTRUCTURAS NO BACILOS	37
3.1 Metodología	37
3.1.1 Adquisición de imágenes	37
3.1.2 Definición de objetos	41
3.1.3 Estandarización de objetos	42
3.2 Extracción y análisis de descriptores de forma	44
3.2.1 Extracción de las características de forma	44
3.2.2 Análisis de las características de forma	44
4 APLICACIÓN DE RECONOCIMIENTO DE PATRONES PARA LA DETECCIÓN DE BACILOS	50
4.1 Introducción	50
4.2 Metodología	50
4.2.1 Criterios de reconocimiento	50
4.2.2 Selección de las características de forma	51
4.3 Resultados	53
CONCLUSIONES	57
IMPLICANCIA	58
RECOMENDACIONES	58
BIBLIOGRAFIA	59

INDICE DE FIGURAS

Figura 1.1: Proceso para diagnosticar paciente con Tuberculosis	11
Figura 1.2: Baciloscopía	13
Figura 1.3: Sistema de detección automática de Tuberculosis	23
Figura 2.1: Imagen microscópica de un bacilo	24
Figura 2.2: Imagen de pixeles de borde del carácter C	27
Figura 2.3: Bacilo segmentado y reconstruido con menos descriptores	29
Figura 2.4: Gráfica de características "a" versus características "b"	31
Figura 2.5: Geometría para la línea de decisión	33
Figura 3.1: Comparativo de muestras sin azul y con azul de metileno	39
Figura 3.2: Diagrama de bloques de la metodología seguida por el autor	40
Figura 3.3: Imagen tomada del microscopio siguiendo el protocolo de adquisición de imágenes establecido por el autor	41
Figura 3.4: Imagen que muestra la segmentación manual de la Figura 3.3	41
Figura 3.5: Imágenes de microscopio y su respectiva segmentación manual	43
Figura 3.6: Espectro de magnitudes promedio de los 350 bacilos	45
Figura 3.7: Espectro de magnitudes promedio de los 350 bacilos en escala logarítmica	45
Figura 3.8: Espectro de magnitudes promedio de los 72 conglomerados	46
Figura 3.9: Espectro de magnitudes promedio de los 72 conglomerados en escala logarítmica	46
Figura 3.10: Espectro de magnitudes promedio de los 58 residuos	47
Figura 3.11: Espectro de magnitudes promedio de los 58 residuos en escala logarítmica	47
Figura 3.12 (a)(b): Comparación de espectro de magnitudes promedio en la zona de bajas frecuencias	49
Figura 4.1: Validación cruzada de 3 capas	52
Figura 4.2: Relación entre cantidad de descriptores y porcentaje de acierto	54

INDICE DE TABLAS

Tabla 4.1: Resultados al ejecutar el programa principal

56



INTRODUCCIÓN

El personal de salud encargado de analizar las muestras de pacientes posiblemente infectados con tuberculosis tiene diversos problemas a la hora de cumplir con su labor. En el Perú las condiciones en la que se encuentran los laboratorios del Ministerio de Salud no son las apropiadas para la seguridad del personal ni para lograr un diagnóstico rápido y confiable, lo que propicia resultados erróneos y aumenta la dificultad de erradicar esta enfermedad.

Los métodos de diagnóstico de la enfermedad más usados en el país son la baciloscopia, cuya deficiencia es la subjetividad del resultado, y el cultivo, que toma mucho tiempo en emitir un diagnóstico. El diagnóstico mediante baciloscopia consiste en la identificación y conteo, por medio de un microscopio, de los bacilos de tuberculosis presentes en una muestra de esputo previamente teñida.

A través del presente trabajo se busca mejorar y agilizar el diagnóstico de tuberculosis en el Perú mediante la aplicación de técnicas de visión artificial. Estrictamente se buscará a establecer las características de forma que permitan discriminar bacilos de otros tipos de artefactos, con la finalidad de posteriormente lograr un algoritmo que automatice la baciloscopía, reduciendo y eliminando sus desventajas actuales, además de contar con las ventajas ya conocidas de la automatización.

El presente documento está organizado en cuatro capítulos. En el primer capítulo se estudia el diagnóstico de tuberculosis, se define el problema, se plantea un sistema completo de detección automática de tuberculosis basado en literatura especializada y se delimita los alcances del presente estudio. En el segundo capítulo se muestra los conceptos teóricos necesarios para el desarrollo de la

propuesta. El tercer capítulo muestra el procedimiento de análisis seguido para seleccionar aquellas características que permitan una óptima discriminación entre las estructuras estudiadas. El cuarto capítulo muestra los resultados de aplicar técnicas de reconocimiento de patrones para diferenciar entre bacilos de Koch y otras estructuras, dando como resultado los descriptores que mejor discriminan entre las clases estudiadas.



CAPÍTULO 1

DIAGNÓSTICO DE TUBERCULOSIS: DEFINICIÓN Y PROBLEMÁTICA

1.1 Introducción

1.1.1 La tuberculosis

La tuberculosis es una enfermedad infecto-contagiosa que viene afectando a la humanidad desde hace mucho tiempo. Existen evidencias paleopatológicas de tuberculosis espinal en esqueletos del neolítico, precolombinos y en restos del antiguo Egipto. Los médicos griegos de la antigüedad usaban el término tisis para indicar su carácter consuntivo. Sin embargo, esta enfermedad no constituyó un problema importante para el ser humano hasta que las condiciones de hacinamiento que establecía la vida urbana de las primeras épocas de la revolución industrial crearon las circunstancias epidemiológicas favorables para su diseminación [1].

Esta enfermedad es causada por diversas especies del genero mycobacterium. La especie más representativa es el bacilo de Koch (***Mycobacterium tuberculosis***). Debido a que la enfermedad se transmite por vía aérea, la bacteria afecta principalmente a los pulmones. Sin embargo, una vez que ingresa al organismo puede producir enfermedad en diferentes órganos como cerebro, riñón, aparato genital, aparato digestivo, etc. [1].

El bacilo de Koch, como algunos otros microorganismos, tiene una pared celular constituida por una gruesa capa de lípidos de la que depende, entre otras propiedades, la ácido-alcohol resistencia. Esta propiedad es la que hace posible el proceso de teñido de bacilos.

Sin embargo, para el estudio, las características que nos interesan son las físicas, ya que son las que se analizarán en el Capítulo 3. De la literatura se sabe que el bacilo de Koch es delgado, ligeramente curvo, con una longitud de 2 a 10 μm y un diámetro de 0.2 a 0.4 μm [2]. El color con el que será teñido el bacilo depende de la técnica de coloración empleada. En el caso de la técnica de Ziehl-Neelsen (ZN) el bacilo se observará de color fucsia.

1.1.2 La tuberculosis en el mundo

Actualmente, la población mundial bordea los 7000 millones de habitantes. Se estima que una tercera parte de la población está infectada con el *Mycobacterium tuberculosis*. Según la Organización Mundial de la Salud (OMS), alrededor de 9 millones de personas desarrolla la enfermedad cada año, de las que se calcula morirán entre 1 a 2 millones. El 95% de los casos se encuentran en los países de bajo y medianos recursos. India y China son los países con más carga de pacientes con tuberculosis a nivel mundial, concentrando el 38% de los casos. [3]

Lamentablemente no existe una vacuna que evite esta enfermedad, por lo que los esfuerzos gubernamentales están dirigidos a la detección temprana de los casos bacilíferos. El objeto de la detección de casos en el control de la tuberculosis es identificar las fuentes de infección en una comunidad, es decir, a las personas que transmiten la infección por el bacilo tuberculoso, para tratarlas y así evitar la transmisión de la enfermedad.

1.1.3 La tuberculosis en el Perú

A lo largo de nuestro territorio, la tuberculosis presenta una distribución irregular, siendo los departamentos con más alta tasa de incidencia Madre de Dios, Lima y Callao, Tacna, Ucayali y Loreto. Esta enfermedad afecta predominantemente a la población económicamente activa más joven, ya que el 52% de este grupo tiene entre 15 y 35 años. Así mismo, el país enfrenta un problema muy grave con respecto al incremento de cepas de *Mycobacterium tuberculosis* que son resistentes y extensamente resistentes a las principales drogas utilizadas en el tratamiento [3][4].

Desde el fortalecimiento del programa de control de la tuberculosis en la década de los 90 hasta al año 2003, se observó una disminución sostenida en la incidencia de TB mayor al 8% anual. Sin embargo, somos el 3er país con más alta tasa de incidencia de América y Caribe, así como también tenemos dificultad en relación al éxito del tratamiento [3][4].

1.1.4 Modo de transmisión

La tuberculosis se transmite por vía respiratoria. La persona con tuberculosis elimina el microbio en las gotitas de saliva al toser, estornudar, escupir, hablar o cantar. Estas secreciones respiratorias dispersas en el ambiente con un número indeterminado de bacilos pueden ser aspiradas de manera frecuente por una persona sana, lo cual ocasiona el contagio [1].

El 3% de las personas infectadas desarrollarán tuberculosis activa durante el año siguiente a la infección, y un 5 a 10% la desarrollará en el transcurso de su vida. [1]

1.1.5 Patogenia

Una vez dentro del ser humano el bacilo de Koch es fagocitado por los macrófagos alveolares. Aproximadamente sólo el 30% de ocasiones los macrófagos son incapaces de destruir al bacilo, produciéndose la infección y generándose un granuloma que histológicamente presenta tejido necrótico. Este granuloma persiste durante toda la vida, sin manifestar síntomas clínicos [1].

1.1.6 Síntomas

Cuando el organismo no logra eliminar al bacilo se desarrolla la enfermedad. La persona que la padece presentará tos y expectoración por más de 15 días como principal síntoma. Además también puede presentar debilidad y cansancio constante, pérdida de peso, fiebre, sudoración nocturna, dolor en el pecho, tos con sangre y pérdida de apetito [1].

1.2 Diagnóstico

El personal de los centros de salud conoce las normas y procedimientos que se deben seguir para detectar pacientes enfermos con tuberculosis. Existen procedimientos para diagnosticar la infección y procedimientos para diagnosticar la enfermedad. La infección es la invasión del organismo huésped por microorganismos, mientras que la enfermedad es el estado alterado de la salud del paciente.

1.2.1 Métodos de diagnóstico

Para el diagnóstico de la infección se busca una respuesta del sistema inmune que nos indique presencia del agente externo, mientras que para el diagnóstico de la enfermedad se requiere de la identificación de la bacteria en el organismo. A continuación una breve reseña de los métodos de diagnósticos más empleados,

1.2.1.1 Métodos de diagnóstico de infección

Reacción de Tuberculina PPD

Actualmente el método más empleado para diagnosticar la infección es la reacción de tuberculina, la cual busca una respuesta inmunológica a la tubérculo-proteínas presentes una vez que el bacilo ha ingresado al organismo. La tubérculo-proteína es un derivado proteico purificado de un extracto de cultivo de bacilo tuberculoso. [5].

La técnica más aplicada para esta prueba es la intradermoreacción de Mantoux, que consta en introducir de manera intradérmica en la cara externa del tercio medio del antebrazo 0,1mL de PPD RT23. Luego de inyectada la solución, se forma una pequeña pápula de 6 a 10mm de diámetro dentro de las siguientes 68 a 72 horas, la cual evidencia infección mas no prueba si hay enfermedad.

Así mismo existen otras técnicas como el Test de Von Pirquet, los parches de tuberculina de Vollmer y el Tine test, que son menos reproducibles.

Medición en sangre de interferón Gamma (IGRAs)

Esta técnica se basa en la medición interferon-gamma, citoquina secretada por los linfocitos T (sensibilizados) cultivados en presencia de dos antígenos (proteínas)

Early Secretory Antigen Target-6 (ESAT-6) y el Culture Filtrate Protein 10 (CFP-10), que son altamente específicas del M. tuberculosis, pero no en las cepas utilizadas en la vacuna BCG. Por este motivo esta técnica es más específica que la prueba en la piel pues no está afectada a vacunas BCG previas. Existen dos test IGRA que se comercializan: el QFT (QuantiFERON TB Gold o QuantiFERON TB Gold in-Tube, laboratorios Cellestis) y el T-SPOT.TB (de laboratorios Oxford Immunotec, Ltd.). El QFT Gold In-Tube incorpora, además de los dos antígenos citados, un tercero, el TB7.7. El QFT mide la concentración de interferón mediante ELISA, mientras el segundo enumera las células T secretoras de interferón. [6]

1.2.1.2 Métodos de diagnóstico de enfermedad

Baciloscopía

Técnica microbiológica que busca teñir la muestra para observar directamente con un microscopio al bacilo de tuberculosis en una muestra de esputo. Esta técnica de diagnóstico se debe a que la pared del bacilo, constituida por lípidos, es ácido-alcohol resistente. Esta propiedad permite que luego de la aplicación del colorante a la muestra (fucsina o auramina) las bacterias queden teñidas y no se decoloren por la acción del ácido-alcohol que se encargará de decolorar los otros elementos de la muestra. La concentración más baja de bacilos que se detecta mediante esta técnica es de 5000 a 10000 por mL de muestra. [5]

Cultivo

Esta técnica se basa en el aislamiento de la micobacteria en diferentes medios preparados, en los que se observa su crecimiento y de acuerdo a esto se diagnostica la enfermedad. El cultivo se puede realizar en medio sólido (huevo coagulado-Lowenstein, agar 7H10 y 7H11 de Middlebrook) o líquidos (bactec MGIT960, BacTALERT 3D, ESPII). Lamentablemente este proceso toma

demasiado tiempo porque la reproducción de los bacilos de tuberculosis es bastante lenta, con tasas de división en el orden de las horas. Esta técnica es la más sensible y específica, pues solo se necesita 10 bacilos para que el resultado sea positivo [5].

La técnica de cultivo radiométrico permite obtener un resultado en aproximadamente 13 días, sin embargo el inconveniente de esta metodología es la manipulación de material radioactivo. Actualmente hay equipos de cultivo radiométrico instalados en algunos laboratorios del país.

Los sistemas no radiométricos también utilizan medios líquidos, con lectura automatizada continua pero con la diferencia que no usan material radioactivo. Estos equipos detectan la bacteria ya sea mediante sistemas fluorescentes que detectan el consumo de O₂ de la bacteria por emisión de luminiscencia, como también con métodos colorimétricos que detectan la producción de CO₂ de la bacteria. En estos sistemas se exige un mínimo de 42 días y un máximo de 56 días para considerarse el resultado como negativo. Es importante mencionar que estas técnicas requieren de equipos costosos por lo que no es común su uso en países de bajos recursos como el nuestro.

Protocolo de detección en el Perú

El proceso para detectar a una persona enferma de tuberculosis empieza con la detección del sintomático respiratorio. Para esto se realizan dos baciloscopías al esputo del paciente, una por día. Si el resultado de alguna de las pruebas es positivo el paciente se someterá a tratamiento. De lo contrario, y de existir una fuerte sospecha de tuberculosis en el paciente, se le solicita una placa radiográfica y se realiza el cultivo de una muestra de esputo. El resultado negativo de la placa

indicará que el paciente no tiene tuberculosis. En el caso del cultivo, si al cabo de 60 días no se encuentran colonias de micobacterias en la muestra, la persona no tiene la enfermedad. El diagrama de flujo de este procedimiento se muestra en la Figura 1.1.

En el Perú la baciloscopia se realiza con la técnica de tñido Ziehl-Neelsen. Esta técnica consiste en obtener una muestra de esputo del sintomático respiratorio, extenderla en una lámina (frotis), esperar 15 minutos, tñirla con fucsina básica, someterla a fuego hasta apreciar tres humos, esperar 5 minutos, echar alcohol al frotis, echar azul de metileno como tinte de contraste, esperar un minuto, lavarla con agua y esperar 5 minutos. Finalmente Se echa una gota de aceite de inmersión al frotis y se coloca en el microscopio óptico binocular para el conteo de bacilos.

Para la lectura de la muestra se requiere un aumento de 1000. La técnica Ziehl-Neelsen requiere mínimo 10000 bacilos por centímetro cúbico de esputo para que sea posible detectarlos, ya que una cantidad menor a esta cifra implicaría que los bacilos se encuentren sumamente dispersos en la muestra. Al mirar por el microscopio se observa un área circular, la cual se denomina campo. Ahí se encuentran bacilos, precipitaciones de tinte, otros residuos tñidos de fucsia (debido a que contienen lípidos), células y otras bacterias tñidas de azul, y el fondo que presenta color azul y regiones blancas. De acuerdo al número de bacilos se sabrá qué tan contagioso es el paciente. El personal de salud debe leer 100 campos, que equivale a 1% del frotis. Si el total de bacilos en dicha cantidad de campos es menor a 9 se informará el número y si es menor o igual a 99 se informará que el resultado del paciente es positivo una cruz. Si el operador encuentra en cada campo más de un bacilo, bastará con que se lea 50 campos y se informará que el resultado es positivo dos cruces. De la misma manera si se contabiliza más de 10 bacilos en cada campo, bastará con que se lea 25 campos y

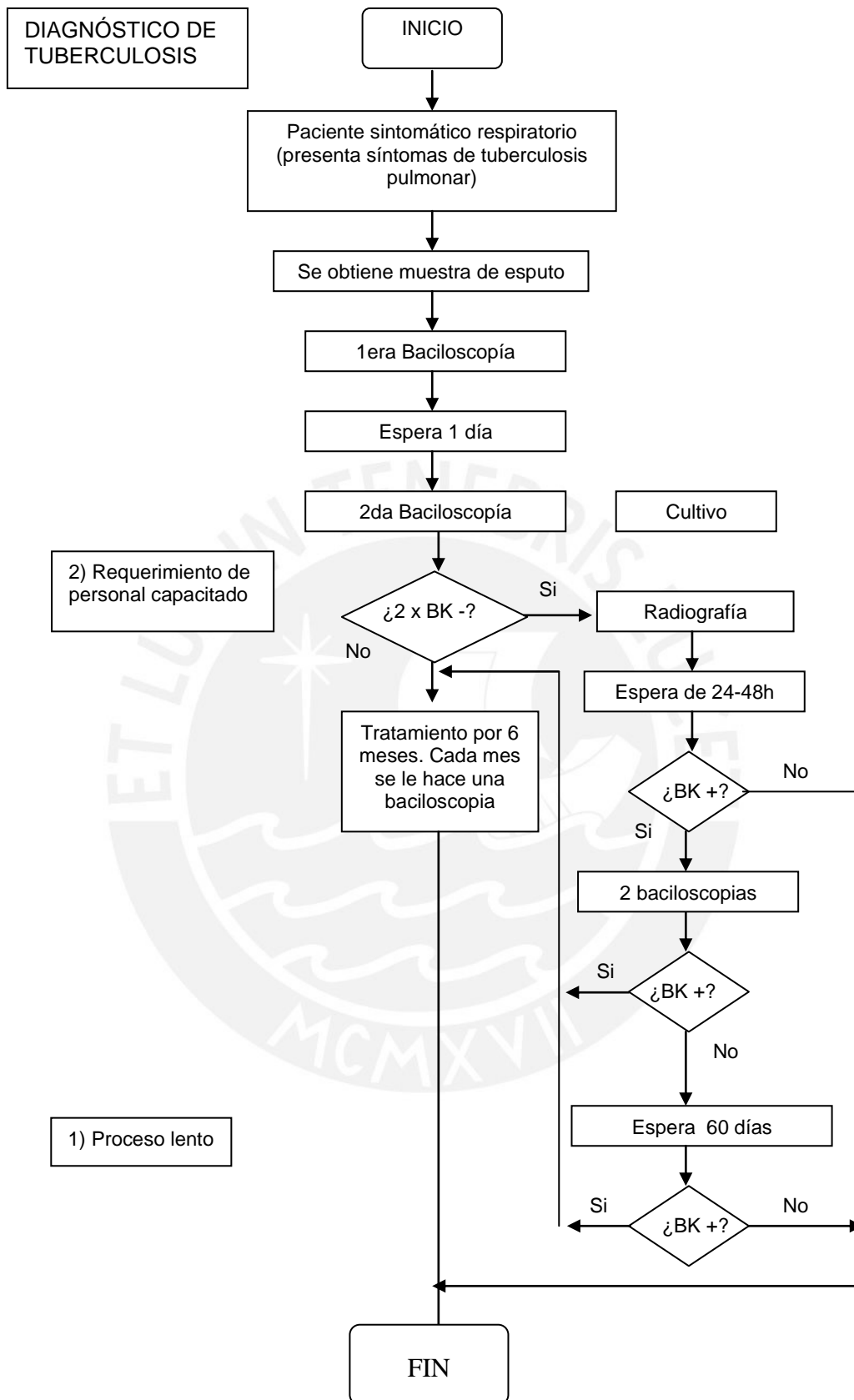


Fig.1.1 Proceso para diagnosticar paciente con tuberculosis

Fuente: Ministerio de salud [7]

se informará que el resultado es positivo 3 cruces. [6][7] El proceso detallado se muestra en la Figura 1.2.

Esta técnica es considerada el examen básico debido a su característica no invasiva y a que se necesita repetirla varias veces para una temprana detección y para controles de tratamiento. La desventaja de esta prueba es que su sensibilidad depende de las características del operador.

1.3 Definición del problema

Esta investigación surge producto de la necesidad de solucionar los problemas detectados en el proceso de diagnóstico mediante baciloscopía, ya que esta técnica es de vital importancia en nuestra realidad. Se distinguen tres factores: el tiempo, el proceso que exige la técnica y el recurso humano.

Se sabe que es crucial la detección rápida de focos de infección, ya que esto permite empezar a tratarlos y evitar que contagien a más personas. Sin embargo, vemos que el primer resultado se emite al día siguiente de tomada la muestra y no es suficiente, por lo que se realiza otra prueba más (Figura 1.1). Por otro lado, el cultivo toma 60 días para emitir un resultado negativo con certeza, por lo que la baciloscopía entrega un resultado de manera más rápida.

Luego de analizar el proceso completo de la técnica de teñido Ziehl-Neelsen, (Figura 1.2), se aprecia que la identificación y conteo de bacilos es la parte más crítica, agotadora y la que más induce al error en los resultados. Leer completamente la muestra le tomaría al personal de salud aproximadamente 17 horas, lo cual es humanamente imposible, por lo que se lee solamente 100 campos, que equivale a 1% del frotis y en algunas ocasiones menos. Debido a que no se lee

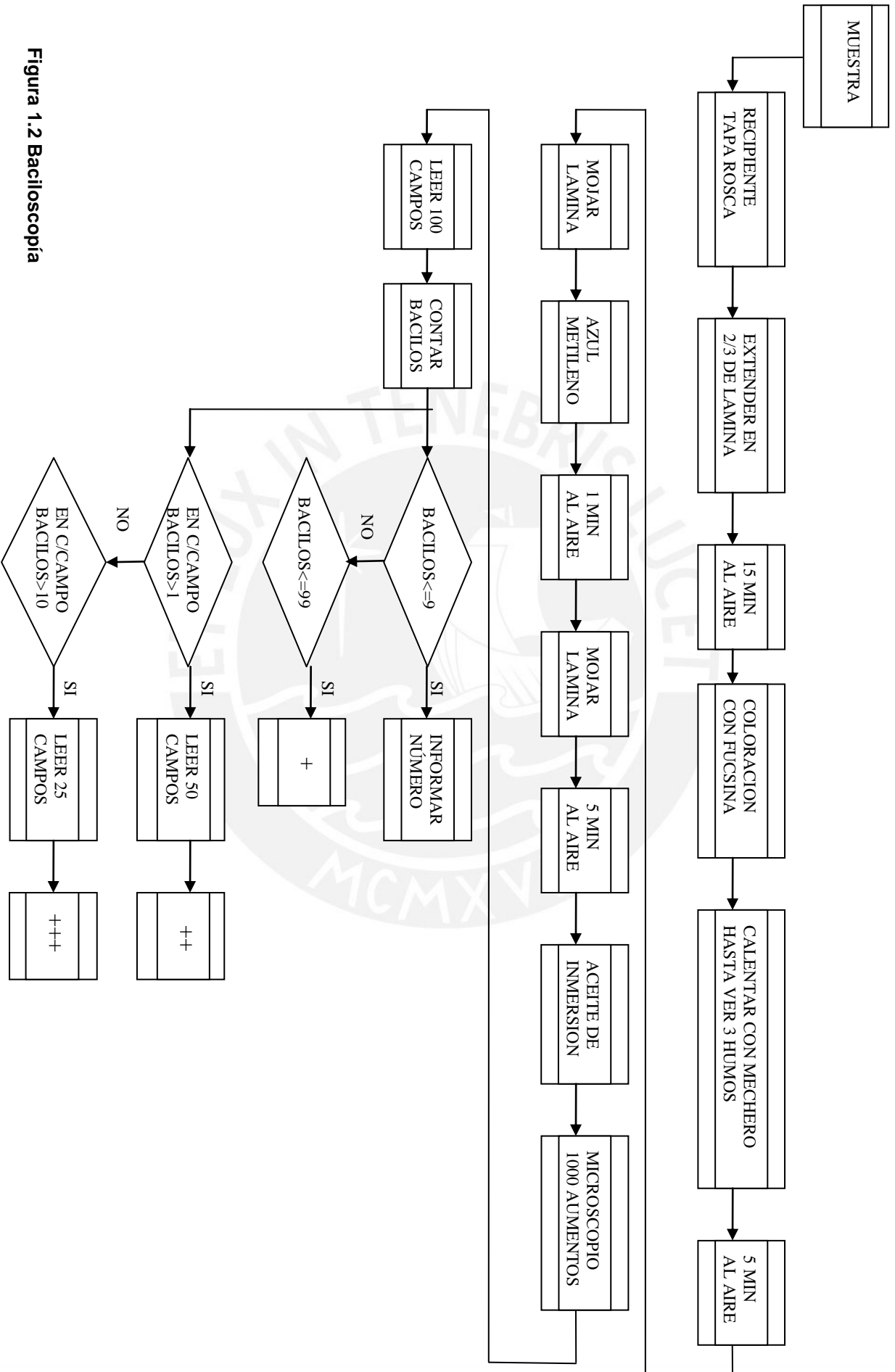


Figura 1.2 Baciloscopía

Fuente: Manual de normas y procedimientos de la baciloscopía para el diagnóstico bacteriológico de la tuberculosis [6]

completamente el frotis se pierde información y se generan falsos resultados, donde los más peligrosos son los falsos negativos.

El otro factor a considerar es la experiencia del personal de salud que realiza las pruebas. El problema está en que la veracidad de los resultados depende de dicha experiencia. Actualmente se busca evitar que los procesos dependan del ser humano, ya que el error es imposible de evitar y de cuantificar.

Debido a que los resultados dependen del operador, se opta por buscar personal calificado, lo que en la realidad peruana es un recurso escaso. Este problema es bastante serio ya que es necesario que el personal encargado tenga conocimiento de normas de bioseguridad, y del procedimiento para el teñido y conteo de bacilos. Adicionalmente el personal debe ser capaz de adquirir una buena muestra, entendiéndose como tal la que proviene del sitio de la lesión que se investiga, obtenida en cantidad suficiente, colocada en el envase adecuado, bien identificada, conservada, y transportada correctamente. Finalmente, el personal debe estar preparado para trabajar durante largas sesiones frente al microscopio evitando que la calidad de los resultados se vea afectada.

1.4 Estado del arte

Con relación al problema planteado y a las posibilidades que brindan los recursos tecnológicos como las técnicas de procesamiento de imágenes, existen diversos trabajos nacionales e internacionales que mediante diferentes técnicas contribuyen en la búsqueda de la automatización de la Baciloscopía. Los diferentes autores concluyen que es posible el desarrollo de un sistema automático de detección de bacilos confiable y fiable.

Konstantinos Veropolous propuso el uso de métodos de aprendizaje automático como redes neuronales y máquinas de vectores de soporte como posibles soluciones de ayuda a la toma de decisiones médicas. [8]

En su investigación contó con dos grupos de muestras de esputo, uno teñido con la técnica de Ziehl-Neelsen y el otro con auramina. El sistema consta de tres partes principales: la captura de imagen, manipulación de la misma y una clasificación objetiva.

La captura de la imagen de frotis, se realizó con ayuda del microscopio. El procesamiento de bajo nivel se desarrolló con las técnicas convencionales de procesamiento de imágenes. Para la detección de borde utilizó el detector de Canny seguido de una umbralización y finalmente una segmentación. Etiquetó cada región con un valor o valores que determinarán lo parecido o diferente de un objeto con respecto al bacilo.

En el entendimiento de alto nivel la imagen entra al proceso de descripción de objetos y extracción de características. Para la selección de características analizó una lista de descriptores de forma y de color. Los descriptores que se consideraron fueron: catorce descriptores de Fourier, compactibilidad, desviación estándar de la intensidad, promedio de intensidad, e intensidad en el centro de masa. Anotó que los momentos invariantes no demostraron propiedades discriminantes. Finalmente para la identificación del bacilo se empleó redes neuronales artificiales y máquinas de vectores de soporte.

Este trabajo fue descrito en dos etapas: primero realiza un estudio preliminar donde muestra un rendimiento de generalización total de hasta 92.1% (alcanzando hasta un 93.5% de sensibilidad y 94.5% de selectividad entre distintos métodos de

clasificación) en un pequeño grupo de 1147 muestras tomadas de cinco pacientes diferentes. Luego trabajó con 65 muestras teñidas con auramina. Los resultados mostraron un rendimiento de generalización ligeramente inferior, ya que alcanzó una precisión global de 87.6% (alcanzando hasta un 93.9% de sensibilidad y un 86,9% de especificidad entre diferentes métodos de clasificación). El autor argumenta que el hecho de considerar las dos técnicas de tinte complicó el análisis de imagen y la clasificación. Así mismo, menciona que es necesario contar con una mayor cantidad de imágenes para que el sistema de clasificación sea considerado estadísticamente factible.

Un trabajo posterior fue el de Forero [9], el cual analiza imágenes de esputo teñidas con auramina. La investigación se centró en la detección, caracterización e identificación de bacilos, los cuales aislaba en sub-imágenes, creadas manualmente desde la imagen original.

La segmentación de imágenes propuesta consta de la detección de borde, operaciones morfológicas y una umbralización especial, que se detalla posteriormente. Para la detección de borde empleó el filtro de Canny seguido de umbralización con histéresis solamente en el canal verde, debido a que la información del bacilo fue abundante en ese canal. Los bordes discontinuos de algunos objetos fueron tratados con operaciones morfológicas de bajo (dilatación y erosión) y alto (apertura y cerrado) nivel. Las imágenes segmentadas contienen todos los objetos que tienen el mismo color del bacilo, sin embargo no todos estos objetos tienen la forma ni el tamaño del bacilo por lo que se les puede eliminar fácilmente mediante un filtro.

Para la caracterización del bacilo se concluyó que los primeros tres y el 11vo momento de Hu fueron suficientes para describir la forma del bacilo, descartando descriptores de Fourier.

Para la etapa de identificación se obtuvo la media y la desviación estándar de cada descriptor, se asumió que la forma del bacilo tiene distribución Gaussiana. Se realizó un análisis de grupos con características similares y para caracterizarlos se usaron los modelos mixtos Gaussianos. Sus resultados arrojaron una sensibilidad mayor a 92% y una especificidad mayor a 97%. Cabe mencionar que los resultados son expresados por campo visual y no por reconocimiento de bacilo, por lo que no es posible una comparación apropiada con otros estudios.

Otro de los estudios que buscó identificar al bacilo en imágenes digitales fue el Sadaphal [17], quien también trabajó con muestras teñidas con la técnica Ziehl-Neelsen. Para la segmentación de color propone una segmentación bayesiana para predecir la probabilidad de que un pixel represente un “objeto TB” usando a priori el conocimiento de los colores de teñido de la técnica Ziehl-Neelsen. Además de un análisis de forma y tamaño. Se señala que los bacilos presentan valores de rojo verde y azul que los diferencia de los objetos no bacilos. Por lo que mediante una umbralización se generó una máscara binaria que fue mejorada usando dilatación morfológica de la imagen.

Para la etapa de extracción de forma, se agruparon los pixeles conectados entre vecinos más cercanos, y se consideraron los descriptores de forma: relación de eje y excentricidad, por ser invariantes a rotación, traslación, inclinación y escala. El autor escogió empíricamente los umbrales para ambos descriptores. Este estudio no muestra los resultados en términos de sensibilidad ni especificidad.

Por otro lado, el trabajo de Sotaquira [18] describe un algoritmo independiente de la iluminación y que además cuenta los bacilos dentro de un conglomerado. Con esta información emitía el diagnóstico basado en la cantidad de bacilos por campo y de acuerdo a la cantidad de campos observados, tal como lo hacen tecnólogos en la práctica. El algoritmo presentó una sensibilidad del 90.9% y una especificidad de 100%. La exactitud lograda es de 85.7% que representa el número de diagnósticos correctos en términos de nivel de infección.

Otro trabajo que se consideró es la tesis de Juan Sato[19], quien presentó un algoritmo automatizado para detectar y contar bacilos de tuberculosis en imágenes de muestras de esputo. Analizó diferentes espacios de color con el fin de hallar el que posea un mayor contraste entre las intensidades de color de los píxeles de bacilos y del fondo. El autor desarrolló una técnica de umbralización adaptativa utilizando el método de Otsu para hallar el óptimo valor umbral. La clasificación la realizó mediante un árbol de clasificación utilizando características de área y excentricidad. La sensibilidad, especificidad y exactitud obtenida estuvieron por encima del 90%.

Otro artículo a considerar es el de Costa et al [20] el cual presenta un método con dos etapas principales: segmentación de bacilo y post-procesamiento. En la segmentación se investigó dos clasificadores: máquinas vectoriales de soporte y redes neuronales. Las variables de entrada a la etapa de segmentación fueron combinaciones de características de color de pixel seleccionadas de 4 espacio de colores (RGB, HSI, YCbCr y Lab), las cuales fueron seleccionadas desde un set inicial de 30 características mediante una técnica de selección escalar. Se produjeron sets con 4,5,6,7 y 8 características. Luego de la segmentación, aún quedan artefactos que deben ser eliminados, para lo cual se aplicaron tres filtros: el primero que eliminó grandes áreas y pequeñas áreas, el segundo que eliminó

objetos de acuerdo a la excentricidad y el tercero un filtro basado en regla que usó el parámetro ratio de color. El estudio arrojó una sensibilidad de 96.8% al identificar bacilos y un error de 3.38%.

1.5 Ventajas de la automatización

La principal ventaja de la automatización es la exactitud en el diagnóstico, ya que se leerá el frotis completo y no el 1% que se lee actualmente. Esto también mejoraría la eficiencia del proceso de exploración al eliminar el cansancio del operario [9].

La automatización del sistema aumentaría el volumen de los análisis, lo cual permitiría realizar baciloscopías de monitoreo a los pacientes que se encuentran en tratamiento y facilitaría la tarea de aislar más rápido a los pacientes multi-drogo resistentes [9].

Por otro lado reduce el riesgo de infección del personal de salud que trabaja directamente con las muestras, así como la contaminación entre frotis [9].

Finalmente gracias a la automatización se puede reducir las horas hombre y el número de personal que analiza las muestras, y reasignar estos recursos a otras tareas necesarias en los laboratorios. [9].

1.6 Sistema de detección automática de tuberculosis

Luego de haber identificado el problema y conocer estudios previos donde se exponen propuestas para la semi-automatización de la baciloscopía se mostrará el

proceso completo para tal propósito, delimitando la parte de la que se ocupa este trabajo.

En la Figura 1.3 se ve el diagrama de bloques de la solución completa propuesta en la tesis de Veropolous [8] para un sistema de detección automática de tuberculosis con imágenes teñidas bajo la técnica Ziehl-Neelsen.

A continuación se explica brevemente los bloques del sistema:

- 1.- Captura de imagen: Se realiza con una cámara digital adaptada a un microscopio.
- 2.- Tratamiento de color: El autor transforma las imágenes a color a imágenes en a escala de grises.
- 3.- Segmentación. En este proceso se busca dividir la imagen en partes que corresponden a un objeto o área. En este caso, la imagen se divide en objetos potencialmente bacilos y fondo. La segmentación se realizó mediante umbralización.
- 4.-Detección de borde.- Este proceso se usa para reducir la información sin perder lo importante, que en este caso es el borde del objeto. Para este paso, el autor propone el detector de Canny [8].
- 5.-Identificación de regiones y filtrado. Luego de la detección de borde cada región es etiquetada mediante un único nivel de gris o color y cuantificada. Las regiones que están fuera de cierto rango o tamaño son filtradas, en este caso basándose en la morfología del bacilo.

6.- Trazado de borde. El borde es necesario para el uso de los descriptores de forma. En este paso se traza el borde interno del objeto y se obtiene la información de la posición de los píxeles del borde. Paso previo a la extracción de características en donde se extrae el borde de la imagen.

7.- Descripción de forma. Este proceso es el central en el presente estudio. Los descriptores de forma son parámetros que dan una representación numérica de la forma de una región.

8.- Clasificación del objeto. Las características obtenidas ingresan a un clasificador, el cual identificará los objetos como bacilos o no bacilos. Se recomienda probar más de un clasificador para poder seleccionar el que mejor desempeño muestre.

El presente trabajo se centrará en los bloques 6, 7 y 8 arriba descritos. Sin embargo, también se describe el proceso de captura de la imagen (bloque 1). Este trabajo de tesis considera objetos manualmente segmentados por el autor de imágenes de baciloscopia de muestras del Laboratorio del Hospital Dos de Mayo. Los objetos segmentados serán procesados para estudiar las características de borde del mismo.

1.7 Objetivos

1.7.1 Objetivo principal

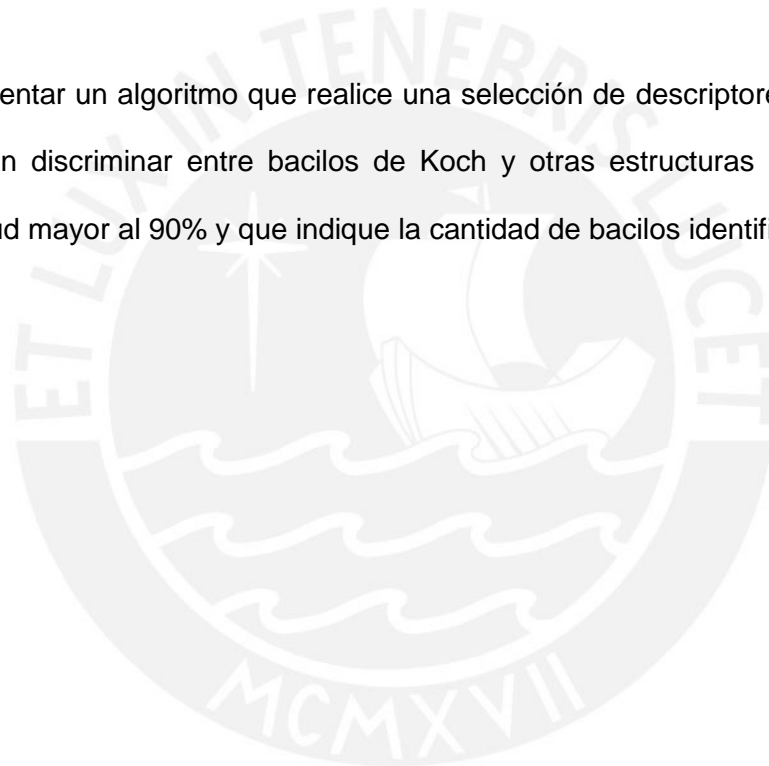
Desarrollar un algoritmo capaz de identificar bacilos de Koch usando procesamiento de imágenes mediante selección de descriptores de Fourier que discriminen de manera óptima entre bacilos y otras estructuras, obtenidas luego de una segmentación manual, con un porcentaje de acierto mayor al 90%.

1.7.2 Objetivos secundarios

Establecer un protocolo de adquisición de imágenes para estandarizar las imágenes captadas.

Implementar un algoritmo que realice la extracción de descriptores de Fourier de estructuras previamente segmentadas.

Implementar un algoritmo que realice una selección de descriptores de Fourier que permitan discriminar entre bacilos de Koch y otras estructuras con un grado de exactitud mayor al 90% y que indique la cantidad de bacilos identificados.



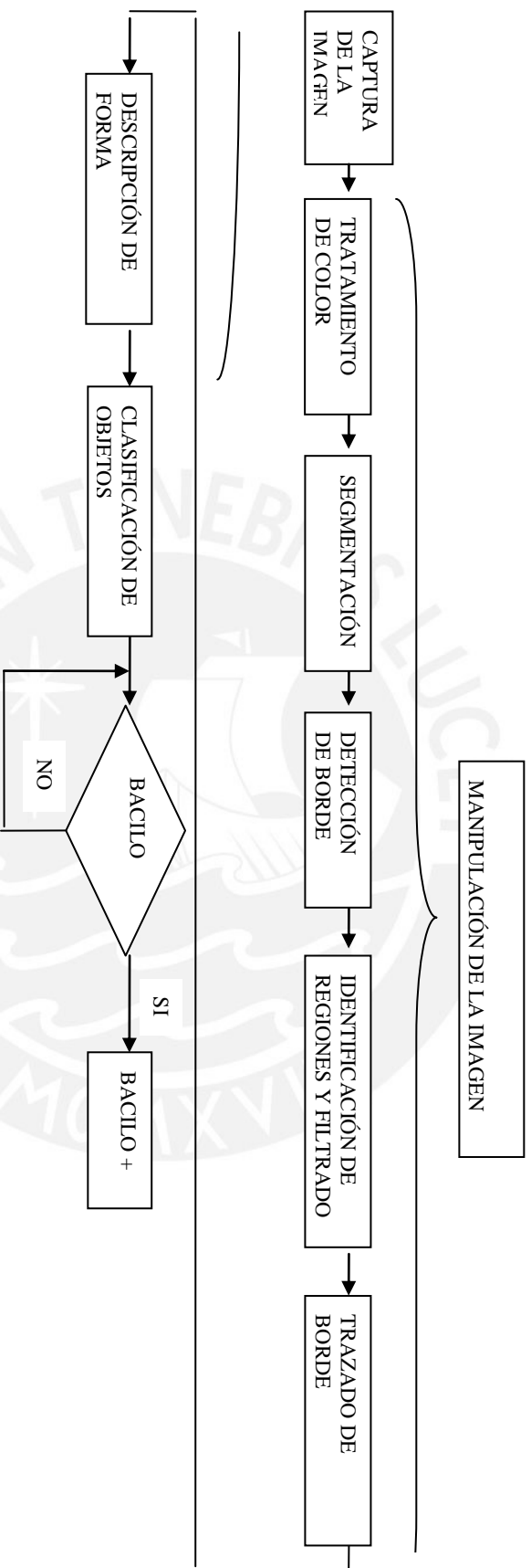


Figura 1.3 Sistema de detección automática de tuberculosis

Fuente: VEROPOULOS, Konstantinos [8]

CAPITULO 2:

CONCEPTOS GENERALES DE DESCRIPTORES DE FORMA

2.1 Modelo teórico

Debido a que este trabajo se orienta a la automatización de un proceso hasta ahora realizado por seres humanos, es necesario desarrollar un sistema que imite la toma de decisión del personal experimentado. Para conseguir este objetivo se aplicarán técnicas de inteligencia artificial [14].

La geometría del objeto en estudio (bacilo) es típicamente un bastón alargado como se observa en la Figura 2.1, mientras que otros objetos pueden presentar distintas formas que van desde una circunferencia hasta formas estrelladas como veremos más adelante. Debido a la geometría de las mencionadas estructuras se optó por trabajar con características de forma, específicamente descriptores de Fourier.



Figura 2.1: Imagen microscópica de un bacilo

2.2 Descriptores de forma

Esta sección se basa en los conceptos presentados por González y Woods [11] [12]. La descripción de objetos es el paso previo y necesario para la interpretación y

entendimiento de las imágenes. El descriptor de forma es un parámetro o grupo de parámetros que mediante una representación numérica describe la forma o borde de una región. El resultado de este paso (descripción) es un descriptor o descriptores y ya no una imagen.

A continuación se presentan algunos descriptores de forma:

Perímetro

Este descriptor sólo tiene significado válido en imágenes binarias. Si consideramos que cada pixel perteneciente al objeto tiene valor 1 y que el resto de pixeles tiene valor cero (fondo), entonces el perímetro viene dado por el número total de pixeles que pertenecen al objeto y que al menos tienen un vecino que pertenece al fondo. La distancia entre pixeles en diagonal se suele tomar como $\sqrt{2}$.

Momentos

Una imagen en escala de grises puede ser únicamente representada por momentos, que interpreten la función imagen como una función probabilística de densidad de variables aleatorias en dos dimensiones. Los momentos pueden ser usados para describir regiones en imágenes binarias o en escala de grises, pero no son invariantes al escalamiento, rotación y traslación, así como también dependen de las transformaciones a escala de grises.

En una imagen digital $f(x,y)$ el momento de orden $p + q$ está dado por:

$$m_{p,q} = \sum_x \sum_y x^p y^q f(x,y) \dots\dots\dots (2.1)$$

Momentos de Hu

Para lograr que los momentos sean invariantes a la traslación se pueden usar los momentos centrales, que en imágenes digitales están dados por:

$$\mu_{pq} = \sum_x \sum_y (x-x_c)^p (y-y_c)^q f(x, y) \dots\dots\dots(2.2)$$

donde

$$x_c = m_{10}/m_{00} \dots\dots\dots(2.3)$$

$$y_c = m_{01}/m_{00} \dots\dots\dots(2.4)$$

x_c y y_c son las coordenadas del centro de gravedad (centroide)

Para que sean invariantes a la escala se pueden usar los momentos centrales escalados

$$n_{pq} = \mu'_{pq}/(\mu'_{00})^\gamma \dots\dots\dots(2.5)$$

y el momento central normalizado sin escalar

$$\theta_{pq} = \mu_{pq}/(\mu_{00})^\gamma \dots\dots\dots(2.6)$$

donde

$$\gamma = ((p+q)/2) + 1 \dots\dots\dots(2.7)$$

$$\mu'_{pq} = (\mu_{pq})/\alpha^{p+q+2} \dots\dots\dots(2.8)$$

(cambio de escala $x'=\alpha x$, $y'=\alpha y$).

La invarianza a la rotación se consigue usando los momentos centrales normalizados sin escalar θ_{pq} , para $p,q=0,1,\dots,3$. Una lista de 7 momentos invariantes se puede derivar de esto:

$$\Phi_1 = \theta_{20} + \theta_{02} \dots\dots\dots(2.9)$$

$$\Phi_2 = (\theta_{20} - \theta_{02})^2 + 4\theta_{11}^2 \dots\dots\dots(2.10)$$

$$\Phi_3 = (\theta_{30} - \theta_{12})^2 + (3\theta_{21} - \theta_{03})^2 \dots\dots\dots(2.11)$$

$$\Phi_4 = (\theta_{30} + \theta_{12})^2 + (\theta_{21} + \theta_{03})^2 \dots\dots\dots(2.12)$$

$$\Phi_5 = (\theta_{30} + 3\theta_{12})(\theta_{30} + \theta_{12})[(\theta_{30} + \theta_{12})^2 - 3(\theta_{21} + \theta_{03})^2] +$$

$$(3\theta_{21} + \theta_{03})(\theta_{21} + \theta_{03})[3(\theta_{30} + \theta_{12})^2 - 3(\theta_{21} + \theta_{03})^2] \dots \dots \dots (2.13)$$

$$\Phi_6 = (\theta_{20} - \theta_{02})[(\theta_{30} + \theta_{12})^2 - (\theta_{21} + \theta_{03})^2] + 4\theta_{11}(\theta_{30} + \theta_{12})(\theta_{21} + \theta_{03}) \dots (2.14)$$

$$\Phi_7 = (3\theta_{21} - \theta_{03})(\theta_{30} + \theta_{12})[(\theta_{30} + \theta_{12})^2 - 3(\theta_{21} + \theta_{03})^2] - (\theta_{30} - 3\theta_{12})(\theta_{21} + \theta_{03}) [3(\theta_{30} + \theta_{12})^2 - (\theta_{21} + \theta_{03})^2] \dots \dots \dots (2.15)$$

Descriptores de Fourier

La Figura 2.2 muestra un borde de N puntos de una imagen digital. Si empezamos en un punto arbitrario (x1,y1), los pares de coordenadas (x1,y1), (x2,y2),..., (xn,yn) representan dicho borde.

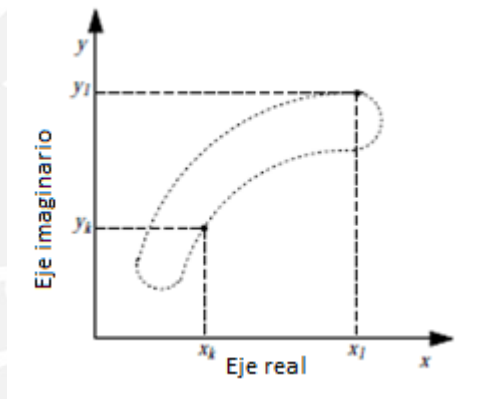


Figura 2.2: Convención de ejes para coeficientes de Fourier. Imagen extraída de Veropolous [8]

Entonces el borde está representado por la secuencia de pares coordenadas $s(k)=[x(k),y(k)]$ para $k=1,2,\dots,N$. Estas coordenadas se pueden expresar como una secuencia de números complejos de la forma:

$$s(k)=x(k)+jy(k) \dots \dots \dots (2.16)$$

para $k=1,2,\dots,N$. Esta transformación permite representar el borde de un objeto bidimensional como una secuencia unidimensional. Luego se aplica transformada discreta de Fourier a (2.16) obteniendo:

$$a(u)=1/N \sum_{k=1}^N s(k)\exp(-j2\pi uk/N) \dots \dots \dots (2.17)$$

para $u=1,2,\dots,K$. Donde $a(u)$ son los coeficientes o descriptores de Fourier. Estos coeficientes son influenciados por la curvatura del objeto y el punto inicial de la curva.

Para que los descriptores de Fourier sean invariantes a la traslación y a la rotación se usa la siguiente transformación:

$$r(u) = \sqrt{|a_x(u)|^2 + |a_y(u)|^2} \dots \dots \dots (2.18)$$

donde $a_x(u)$ y $a_y(u)$ son el resultado de la transformada de Fourier de la parte real y la parte imaginaria respectivamente del número complejo $s(k)$.

Para que sean invariantes ante cambios en la escala se usa la transformación:

$$w(u) = r(u) / r(1) \dots \dots \dots (2.19)$$

Los primeros coeficientes (bajas frecuencias) determinan la forma global del contorno del objeto, mientras que los últimos coeficientes (altas frecuencias) son los responsables de los detalles más finos. Se puede obtener la magnitud de estos coeficientes y obtener un espectro de frecuencias. Lógicamente las magnitudes serán mayores mientras mayor información contenga el descriptor. Los objetos simples (i.e. sin mucho detalle ni curvas pronunciadas) tendrán la información en las bajas frecuencias, mientras que objetos más elaborados tendrán un mayor ancho de banda.

Estos descriptores pueden generar nuevamente el contorno del objeto mediante la transformada inversa de Fourier, la cual dará como resultado la secuencia de números complejos vista en (2.16).

El espectro de magnitudes de los descriptores de Fourier muestra que la información está concentrada en cierta porción del espectro, es decir, hay descriptores con información útil y descriptores con información de la que se puede

prescindir. La importancia de esto radica en que se puede reconstruir una imagen a partir de un grupo reducido de descriptores. La cantidad de descriptores a usar para la reconstrucción depende de la complejidad de la curva e influye directamente en el borde.

La Figura 2.3 muestra el borde de un bacilo, que se obtuvo al aplicar segmentación basada en detección de borde, junto con una secuencia de la misma imagen a la cual se le representa con menos descriptores de Fourier.

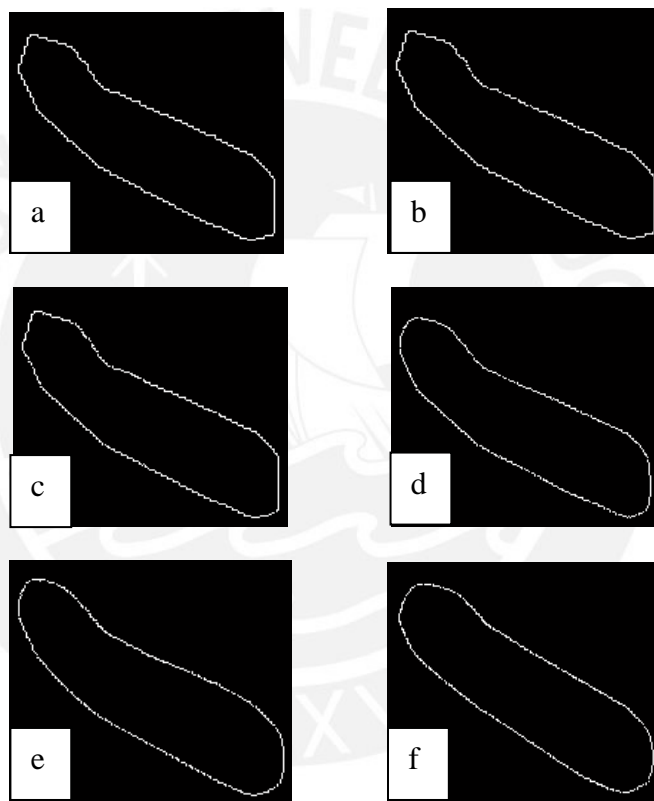


Figura 2.3 Bacilo segmentado y reconstruido con menos descriptores. a) Muestra el bacilo segmentado mediante algoritmo basado en detección de borde. b) Bacilo reconstruido con 80 descriptores. c) Bacilo reconstruido con 40 descriptores. d) Reconstrucción con 20 descriptores. e) Reconstrucción con 10 descriptores. f) Reconstrucción con 5 descriptores.

2.3 Reconocimiento de patrones

Un mayor detalle de los conceptos explicados en esta sección se puede encontrar en [10] y [11]. El reconocimiento de patrones es la disciplina científica que busca clasificar objetos (incluyendo imágenes) en categorías o clases.

La clasificación requiere identificar las características medibles (patrones) que hacen la diferencia entre dos o más tipos/clases de objetos. En la práctica, los dos principales arreglos de patrones son vectores (para descripciones cuantitativas) y cadenas (para descripciones estructurales). Los vectores de patrones tienen la forma $n \times 1$.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \dots\dots\dots (2.20)$$

donde la componente x_i , representa el i -ésimo descriptor y n es el número total de los descriptores asociados al patrón. El reconocimiento de patrones usando vectores cuantitativos se basa en el cálculo de distancias, donde la distancia se define como un número que cuantifica el grado de similitud entre dos objetos. Se basa además en el uso de funciones de decisión o discriminantes.

Para un mejor entendimiento se considera el siguiente caso. Se asume que los objetos pueden pertenecer a dos clases, A y B, y que cada objeto cuenta con dos características o descriptores, a y b, que sirven para la clasificación. En la Figura 2.4, se observan círculos y triángulos que corresponden a un objeto diferente. Se aprecia que los objetos de la clase A se encuentran en una región del plano a-b diferente a los de la clase B. Una manera de dividir las regiones del espacio ocupadas por los objetos de tipos A y B es usar una línea recta, conocida como

línea de decisión, que en nuestro ejemplo sería la línea sólida oblicua. Consideremos un objeto cuya clase se desconoce y lo representamos por una X en el plano a-b. De acuerdo a sus características es ubicado a la izquierda de la línea de decisión, lo que indica que es más probable que pertenezca a la clase A. En este ejemplo, la línea de decisión se usa como un clasificador que divide el espacio de características en regiones correspondientes a cada clase. Los objetos que se usaron para diseñar el clasificador (en este ejemplo los círculos y triángulos) son conocidos como objetos de entrenamiento.

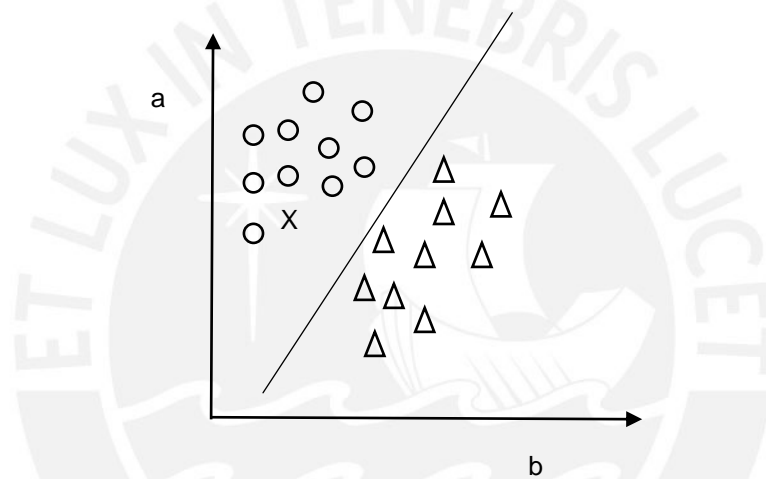


Figura 2.4 Gráfica de característica "a" versus característica "b" de los diferentes objetos de clase A (○) y clase B (Δ). Nuevo objeto a clasificar "X".

Existen diferentes tipos de clasificadores que son utilizados de acuerdo a la aplicación, complejidad y criterio deseado. A continuación una breve explicación de algunos de los clasificadores más importantes y que de alguna u otra forma se estudiaron para desarrollar el presente estudio.

Clasificador basado en la teoría de decisión de Bayes

Consideremos un caso donde existen sólo dos clases de objetos, w_1 y w_2 . Asumimos que se conocen las probabilidades a priori $P(w_1)$ y $P(w_2)$ y que se

conoce la función de densidad de probabilidad condicional $p(\mathbf{x}|w_i)$, $i=1,2$, que describe la distribución de vectores de características en cada una de las clases.

$$P(w_i|\mathbf{x})=[p(\mathbf{x}|w_i)P(w_i)]/p(\mathbf{x}) \dots\dots\dots(2.21)$$

Donde $p(\mathbf{x})$ es la función de densidad de probabilidad de x

$$p(\mathbf{x})= \sum_{i=1:2} p(\mathbf{x}|w_i) P(w_i) \dots\dots\dots(2.22)$$

Luego, la regla de clasificación de Bayes es

$$\text{Si } P(w_1|\mathbf{x}) > P(w_2|\mathbf{x}), \mathbf{x} \text{ es clasificada en } w_1 \dots\dots\dots(2.23)$$

$$\text{Si } P(w_1|\mathbf{x}) < P(w_2|\mathbf{x}), \mathbf{x} \text{ es clasificada en } w_2 \dots\dots\dots(2.24)$$

En el caso de igualdad, el patrón puede ser asignado a cualquiera de las dos clases.

Clasificador basado en análisis discriminante lineal

El objetivo es encontrar la combinación lineal de las variables que permita diferenciar entre clases y luego emplear dicha combinación lineal para predecir a qué grupo pertenecería un nuevo caso.

La función discriminante que es una combinación lineal de los componentes de x puede ser escrita así:

$$g(x) = w^T x + w_0 \dots\dots\dots(2.25)$$

donde $w = [w_1, w_2, \dots, w_i]^T$ es el vector de pesos y w_0 el umbral. Para simplificar consideramos el caso de dos clases y definimos la superficie de decisión igualando la ecuación 2.25 a cero

$$g(x) = w^T x + w_0 = 0 \dots\dots\dots(2.26)$$

Si x_1, x_2 son dos puntos en el hiperplano de decisión, luego lo siguiente es válido:

$$w^T (x_1 - x_2) = 0 \dots\dots\dots(2.27)$$

Ya que el vector $x_1 - x_2$ está dentro del hiperplano de decisión (para cualquier x_1, x_2) y considerando la ecuación 2.27 se tiene que el vector w es ortogonal al hiperplano de decisión.

La Figura 2.5 muestra la geometría correspondiente (para $w_1 > 0$, $w_2 > 0$, $w_0 < 0$).

Se observa que las cantidades ingresadas en la figura están dadas por:

$$d = |w_0| / \sqrt{w_1^2 + w_2^2} \dots\dots\dots(2.28)$$

y

$$z = |g(x)| / \sqrt{w_1^2 + w_2^2} \dots\dots\dots(2.29)$$

En otras palabras, $|g(x)|$ es la medida de la distancia Euclídeana del punto x desde el hiperplano de decisión. A un lado del plano $g(x)$ toma valores positivos y al otro lado negativos. En el caso especial que $w_0 = 0$, el hiperplano pasa por el origen.

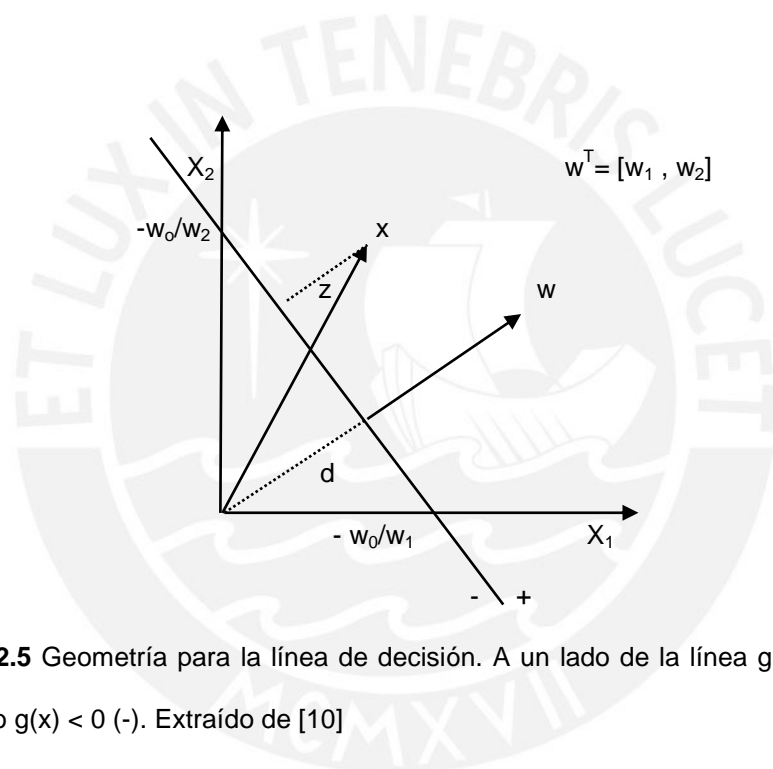


Figura 2.5 Geometría para la línea de decisión. A un lado de la línea $g(x) > 0$ (+) y en el otro lado $g(x) < 0$ (-). Extraído de [10]

2.3.1 Selección de características

Existe más de una razón para reducir el número de características al mínimo. El objetivo es seleccionar las características más importantes buscando retener la información discriminante. Una opción es analizar las características individualmente y descartar aquellas con baja capacidad discriminatoria. La otra opción, y en la que nos enfocaremos en el presente trabajo, es examinar las características en combinaciones [10].

Dentro de la selección del vector de características, de acuerdo con la regla de optimización que se escoja tenemos dos categorías: enfoque de filtro, en la que se selecciona las características en forma independiente del clasificador y el enfoque encapsulado, donde selecciona los subconjuntos en función del desempeño de un clasificador. Para ambos casos, con el fin de reducir la complejidad, existen técnicas de búsqueda subóptima y óptima (bajo ciertas consideraciones) [10].

Técnicas de búsqueda sub óptima

Selección secuencial hacia atrás

Se demostrará el método mediante un ejemplo. Consideramos que se evalúan cuatro características ($m = 4$) : x_1, x_2, x_3, x_4 . Se desea seleccionar las dos que mejor discriminen. El proceso de selección consiste en:

1. Adoptar un criterio de separación, C , y procesar dicho criterio para el vector $[x_1, x_2, x_3, x_4]^T$.
2. Eliminar una característica y por cada una las posibles combinaciones resultantes ($[x_1, x_2, x_3]^T$ $[x_1, x_2, x_4]^T$ $[x_1, x_3, x_4]^T$ $[x_2, x_3, x_4]^T$), se procesa bajo el criterio adoptado. Se selecciona la combinación con mejor resultado.
3. Se aplica el mismo procedimiento para el nuevo vector que contiene 3 características hasta lograr contar con el vector con las 2 características que mejor discriminan.

Este es un procedimiento sub óptimo ya que no se puede garantizar que la óptima combinación de 2 características se origina de un óptimo grupo de 3 características.

[10]

Selección secuencial hacia adelante

Este proceso es la reversa del visto anteriormente. Consideramos que se evalúan las mismas cuatro características ($m = 4$) : x_1 , x_2 , x_3 , x_4 y que se desea seleccionar las dos que mejor discriminen.

1. Se procesa cada característica con el criterio de separabilidad adoptado. Se selecciona la que da el mejor resultado, por ejemplo x_1
2. Se procesa todas las combinaciones posibles de 2 características que contenga al ganador del paso previo, es decir, $[x_1, x_2]^T$, $[x_1, x_3]^T$, $[x_1, x_4]^T$. Se procesa bajo el criterio adoptado y el que ofrezca el mejor resultado es la combinación con las características que mejor discriminan.[10]

Para el presente estudio se optó por la técnica de selección secuencial hacia atrás ya que desde el punto de vista computacional es más eficiente que la otra opción.

Validación de resultados

Luego de obtener resultados de las diferentes técnicas de extracción y clasificación de características, se necesita validar los resultados. Una técnica frecuentemente usada para dicho fin es la validación cruzada de k capas, la cual se basa en un muestreo repetitivo. Así, la muestra se divide en k grupos de tamaño lo más uniforme posible. El clasificador es probado k veces, omitiendo en cada prueba un grupo de la muestra cuyos objetos son usados como datos de prueba. El resultado final se obtiene promediando los resultados obtenidos en cada prueba.

Confiabilidad de resultados

Todo resultado cae dentro de una de cuatro opciones posibles: Verdadero Positivo (VP) y Verdadero Negativo (VN), cuando el sistema predice correctamente la clase

del objeto, Falso Positivo (FP) cuando el sistema clasifica como positivo un objeto que realmente es negativo y Falso Negativo (FN) cuando el sistema clasifica como negativo un objeto que realmente es positivo. La sensibilidad, especificidad y la exactitud se definen a partir de estas cuatro cantidades como:

$$\text{Sensibilidad} = VP / (VP + FN) \dots\dots\dots(2.33)$$

$$\text{Especificidad} = VN / (VN + FP) \dots\dots\dots(2.34)$$

$$\text{Exactitud} = (VP + VN) / (VN + VP + FN + FP) \dots\dots(2.35)$$



CAPITULO 3:
EXTRACCIÓN Y ANÁLISIS DE DESCRIPTORES DE FORMA DE BACILOS Y DE
ESTRUCTURAS NO BACILOS

3.1 Metodología

En este estudio se usó el programa ImageJ (NIH, Maryland, EEUU) para realizar un pre-procesamiento manual de las imágenes digitales obtenidas y el Image Processing Toolbox, del MATLAB (Mathworks, Massachussets, EEUU) versión 7.8, con el cual se realizaron todos los programas necesarios para el estudio.

3.1.1 Adquisición de imágenes

La adquisición de imágenes se realizó en el Laboratorio Central de Hospital Nacional Dos de Mayo, con la supervisión permanente de personal que labora en dicho nosocomio.

Se estableció el siguiente protocolo de adquisición de imágenes:

Protocolo de adquisición de imágenes

1. El proceso de adquisición de imágenes se realizó en el ambiente del laboratorio del Hospital Dos de Mayo medianamente iluminado con luz natural.
2. Se utilizó un microscopio marca Olympus modelo CX31, con un objetivo de 100x y ocular de 10x para la visualización de los microorganismos.
3. Se utilizó una cámara marca Olympus modelo SP 310 de 7 mega píxeles, a la que se le acopló un adaptador para mayor estabilidad, para la adquisición de imágenes.

4. La cámara se posicionó perpendicular y en contacto con la lente del microscopio para evitar el brillo del exterior.
5. Se tomaron las fotografías sin flash y utilizando la opción P de la cámara, la cual controla de forma automática la abertura y velocidad de obturación.
6. Se consideraron muestras de pacientes sin haber recibido tratamiento, de lo contrario los bacilos saldrían divididos.
7. Se debe realizar correctamente el proceso de tinción, extender de manera correcta la muestra en la lámina, deben haber proporciones exactas de las sustancias requeridas y un adecuado control de tiempo de exposición al fuego.
[7]
8. Se debe tomar las imágenes al borde de la muestra para evitar azules muy intensos que perjudican el procesamiento de imágenes.
9. Se debe tomar imágenes donde los bacilos estén completos y no divididos por el límite de la foto.
10. El conteo de bacilos debe ser realizado o en su defecto validado por personal de salud con experiencia en el conteo de bacilos.

Se obtuvo imágenes correspondientes a tres pacientes y 100 tomas por paciente para un total de 300 fotografías. Dichas fotografías fueron divididas en imágenes con más de tres bacilos, imágenes con mínimo un bacilo, e imágenes sin bacilos contables. En total para el presente estudio el autor etiquetó 350 bacilos y 130 objetos no bacilos.

Cabe mencionar que durante el proceso de adquisición de imágenes se pidió al personal de salud teñir una muestra sólo con el color fucsia y prescindir del azul de metileno para obtener mayor claridad, ya que se observó que el azul de metileno muchas veces opaca a los bacilos y el operador invierte tiempo buscando una zona con menos azul. En la Figura 3.1 se hace una comparación entre una foto sin azul y

una con azul. En un posterior estudio se analizará si estas imágenes facilitan la visión artificial o sólo facilita la lectura del operador.



Figura 3.1. Comparativo de muestras sin azul y con azul de metileno. La imagen de la izquierda es una muestra sin azul de metileno. La imagen de la derecha es de una muestra con azul de metileno (muestra común).

El procedimiento seguido en el presente estudio se visualiza en la Figura 3.2. Las imágenes obtenidas del microscopio (ver ejemplo en la Figura 3.3) constan de diversas estructuras, las cuales fueron identificadas bajo la supervisión de personal de salud experimentado. Cada objeto presente en las imágenes es segmentado manualmente con herramientas del software ImageJ (NIH, Maryland, EEUU). En la Figura 3.4 vemos la segmentación de las estructuras de la imagen de la Figura 3.3. Finalmente cada objeto es almacenado individualmente y etiquetado formando una base de datos de 480 objetos.

Por consideraciones de software todas las estructuras de la base de datos se uniformizaron a 317 píxeles, luego se extrajeron y normalizaron sus descriptores de Fourier. Finalmente, se realizó la búsqueda sub-óptima selección secuencial hacia atrás de los 5, 7 y 14 descriptores que mejor discriminan entre las clases de estructuras, es decir entre bacilos y no bacilos.

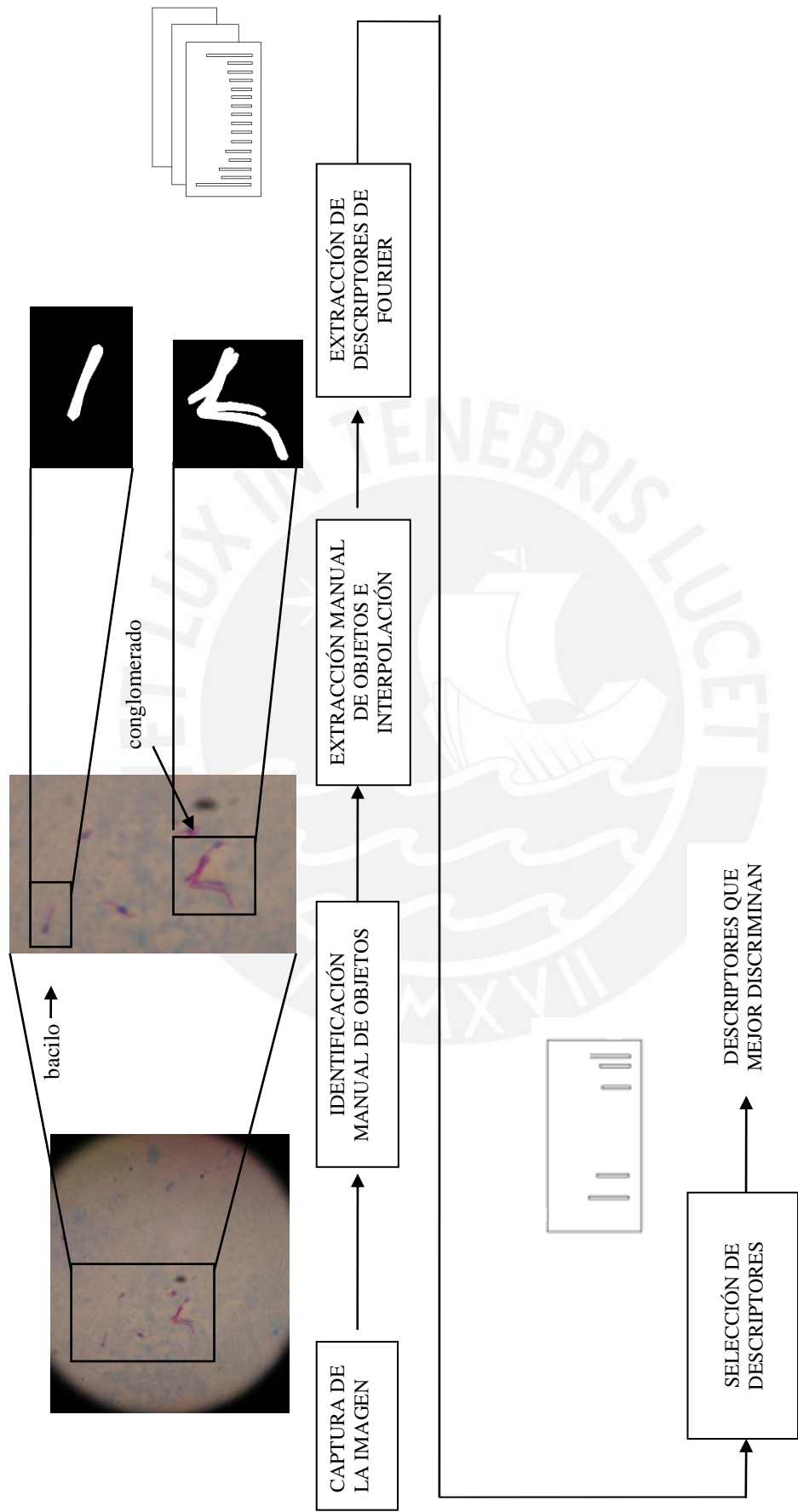


Figura 3.2 Diagrama de bloques de la metodología seguida por el autor.

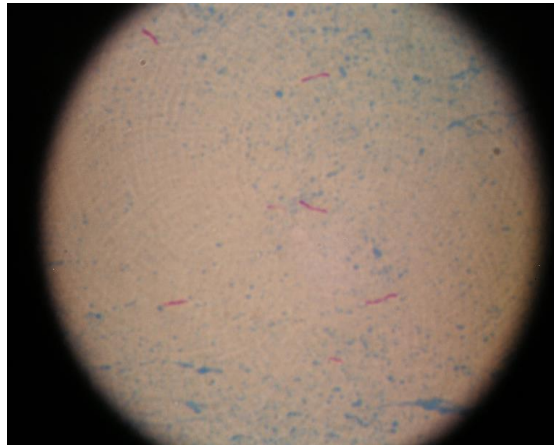


Figura 3.3. Imagen tomada del microscopio siguiendo el protocolo de adquisición de imágenes establecido por el autor

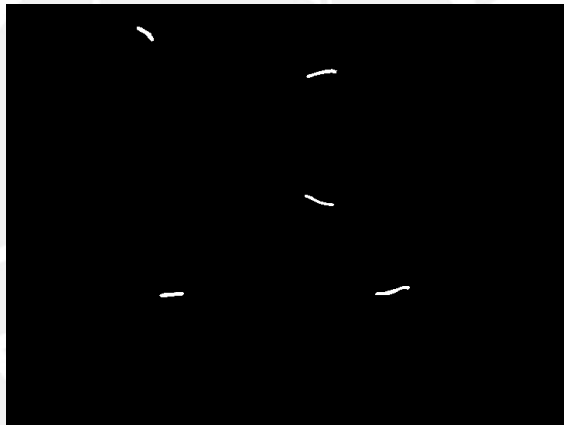


Figura 3.4. Imagen que muestra la segmentación manual de la Figura 3.3

3.1.2 Definición de objetos

Debido a la posición relativa en que se encuentra el bacilo al momento de capturar la imagen, éste muestra variados tamaños y formas. Cabe mencionar que el bacilo se encuentra en un espacio de tres dimensiones, y la imagen que obtenemos es de dos dimensiones. Entonces, si el bacilo se encuentra perpendicular al plano de la foto se verá como una pequeña mancha fucsia ovalada, la cual es similar en forma a las precipitaciones de tinte y/o gérmenes que debido a una mala decoloración o a

que contienen lípidos presentan el color fucsia característico (Ver Figura 3.5). Todos los objetos de esta forma serán llamados residuos.

De la misma manera, otro tipo de objetos de color fucsia son los mismos bacilos que están muy juntos y/o cruzados, a los que se les llamará conglomerado de bacilos. Debido a que el presente estudio busca adoptar el protocolo actualmente aceptado para la baciloscopía, es que nuestro algoritmo solo considera bacilos contables aquellos cuyo borde está bien definido y que presenten la forma característica de bastón alargado [1].

Ni los residuos ni los conglomerados son tomados en cuenta a la hora del conteo manual que realizan los tecnólogos, por lo que el sistema propuesto mantendrá la misma política, principalmente a fin de evitar falsos positivos que indiquen tratar al paciente.

3.1.3 Estandarización de objetos

La base de datos se creó a partir de las 300 imágenes capturadas, de las cuales se segmentó 350 bacilos, 72 conglomerados de bacilos y 58 residuos, es decir 480 objetos. Luego de la etapa de extracción manual, se aprecia que los objetos muestran diferentes tamaños, por lo que se decidió estandarizar las imágenes a un mismo número de píxeles, tal como se mencionó anteriormente a 317 píxeles.

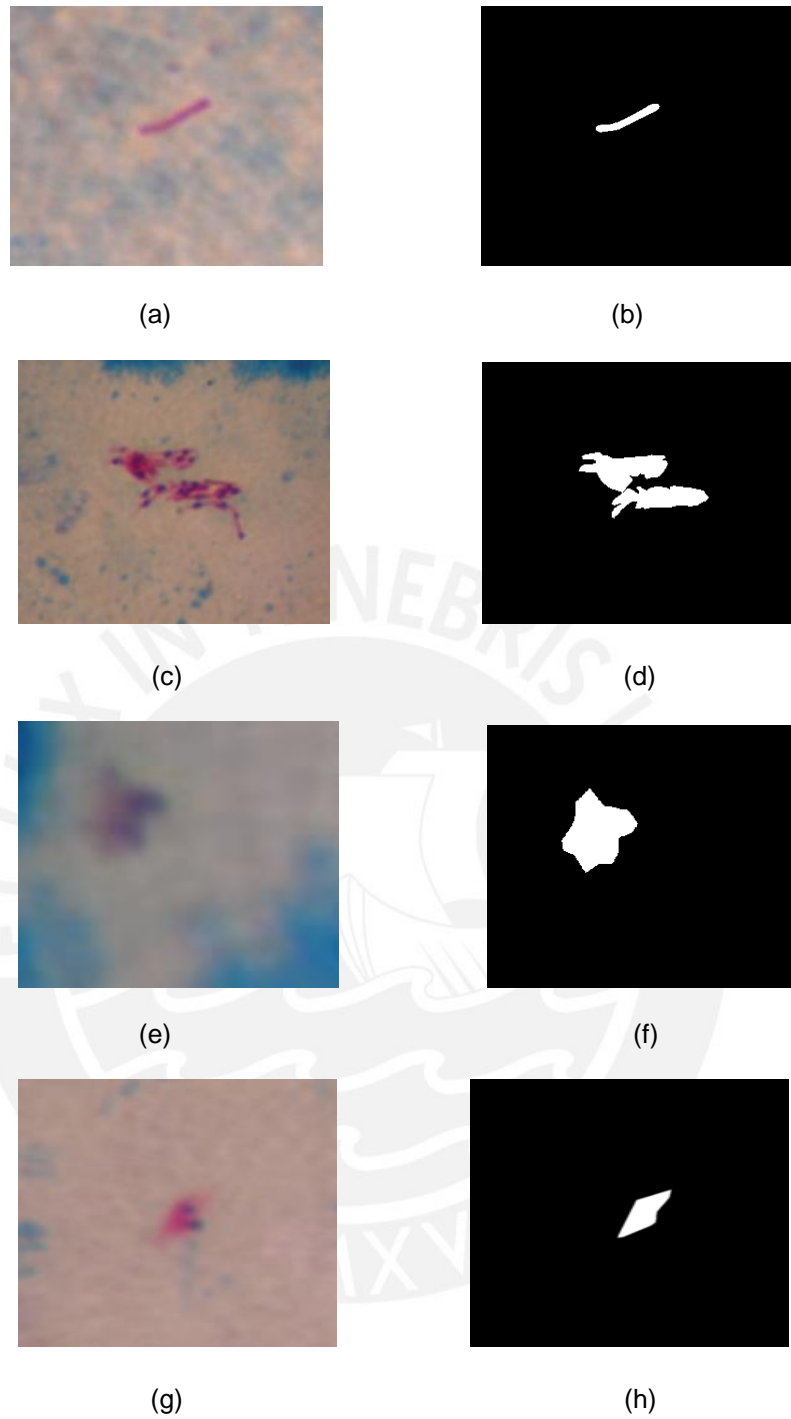


Figura 3.5 Imágenes de microscopio y su respectiva segmentación manual. (a) Bacilo. (b) Bacilo segmentado. (c) Conglomerado de bacilos (d) Conglomerado de bacilos segmentado. (e) Precipitación de tinte. (f) Segmentación de precipitación de tinte. (g) Residuo. (h) Segmentación de residuo.

3.2 Extracción y análisis de descriptores de forma

En esta sección se procede a extraer los descriptores de forma que representan a los tres tipos de objetos a fin de analizar el comportamiento de la información entre ellos. El análisis de dicho comportamiento en bacilos y no bacilos permitirá establecer criterios y consideraciones que finalmente aportarán para la etapa de selección, que buscará identificar a los descriptores que mejor discriminan entre las clases.

3.2.1 Extracción de las características de forma

Cada uno de los 480 objetos fue representado por sus descriptores de Fourier, los cuales se normalizaron con el fin de que sean descriptores invariantes a escala, traslación y rotación. Así, cada objeto quedó caracterizado por 317 descriptores de Fourier.

3.2.2 Análisis de las características de forma

Como era de esperar, el espectro de los bacilos muestra que la mayor cantidad de información se encuentra en las bajas frecuencias, ya que su forma no se caracteriza por tener curvas pronunciadas, tal como se aprecia en las figuras 3.6 y 3.7. Lo que se muestra en la figura 3.6 es el promedio del espectro de magnitudes de los 317 descriptores de Fourier de los 350 bacilos. La figura 3.7 muestra el mismo promedio pero con las magnitudes en escala logarítmica. Debido a que los Descriptores de Fourier de los objetos están normalizados, la magnitud del Descriptor 1 es cero y la del Descriptor 317 es 1. En el caso de los bacilos el promedio de las magnitudes de los descriptores restantes (del 2do al 316) es 0.0039 con una desviación estándar promedio de 0.0372. Para los conglomerados

el promedio de las magnitudes de los descriptores restantes es 0.0062 con una desviación estándar promedio de 0.0332 y para los residuos el promedio para el mismo grupo de descriptores es de 0.0024 con una desviación estándar promedio de 0.0141. Notamos que la media y la desviación estándar no aportan gran ayuda al momento de discriminar ya que los rangos donde se encuentran estos valores en las diferentes clases de objetos se traslapan.

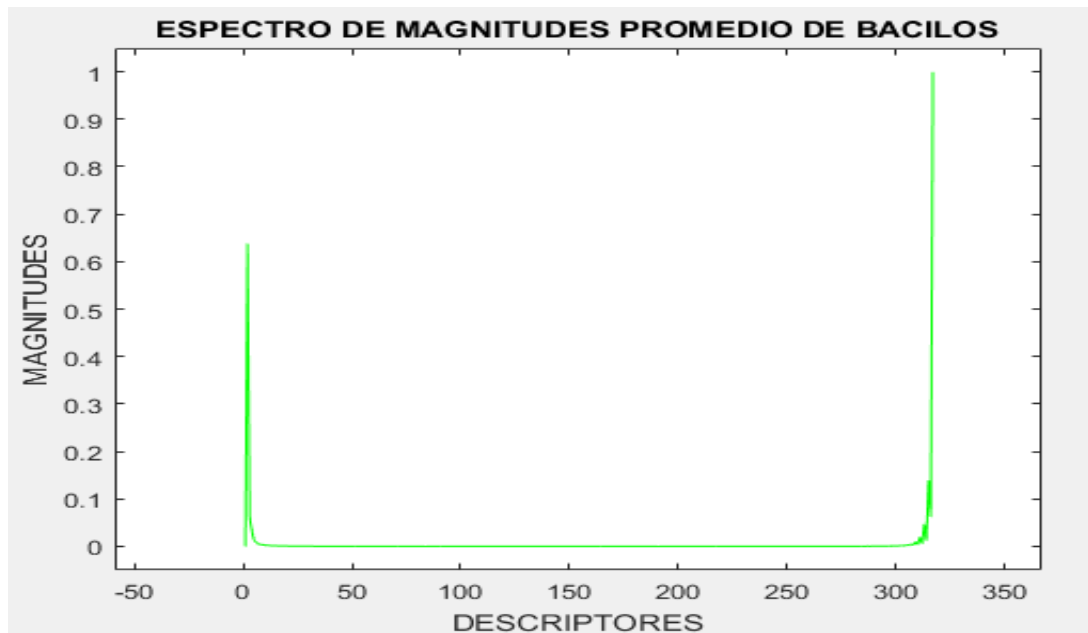


Figura 3.6. Espectro de magnitudes promedio de los 350 bacilos.

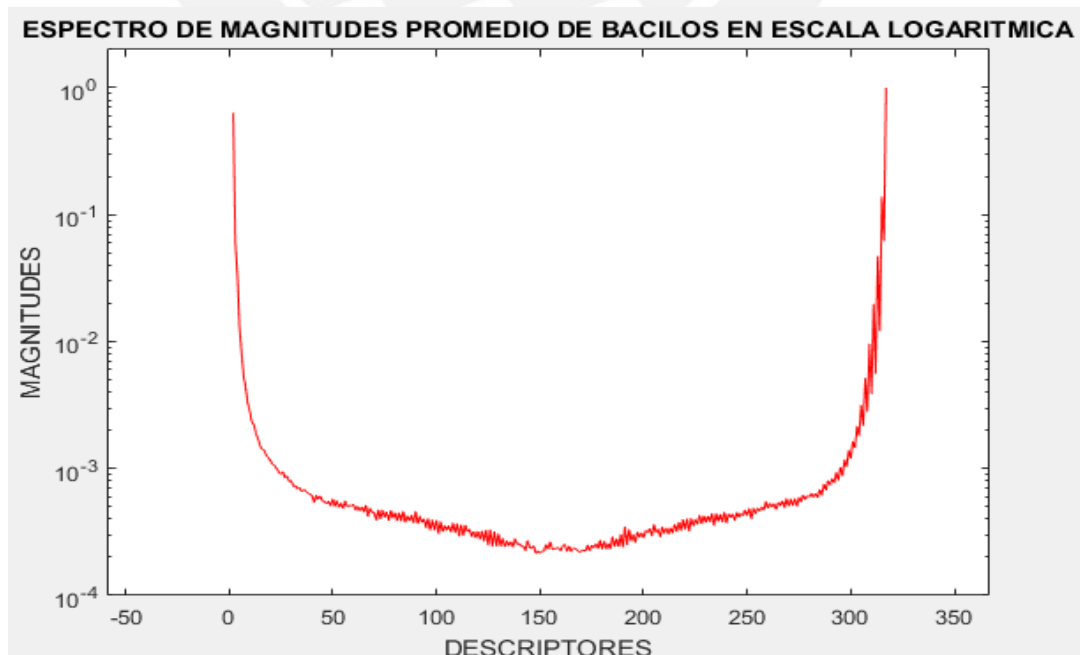


Figura 3.7. Espectro de magnitudes promedio de los 350 bacilos en escala logarítmica.

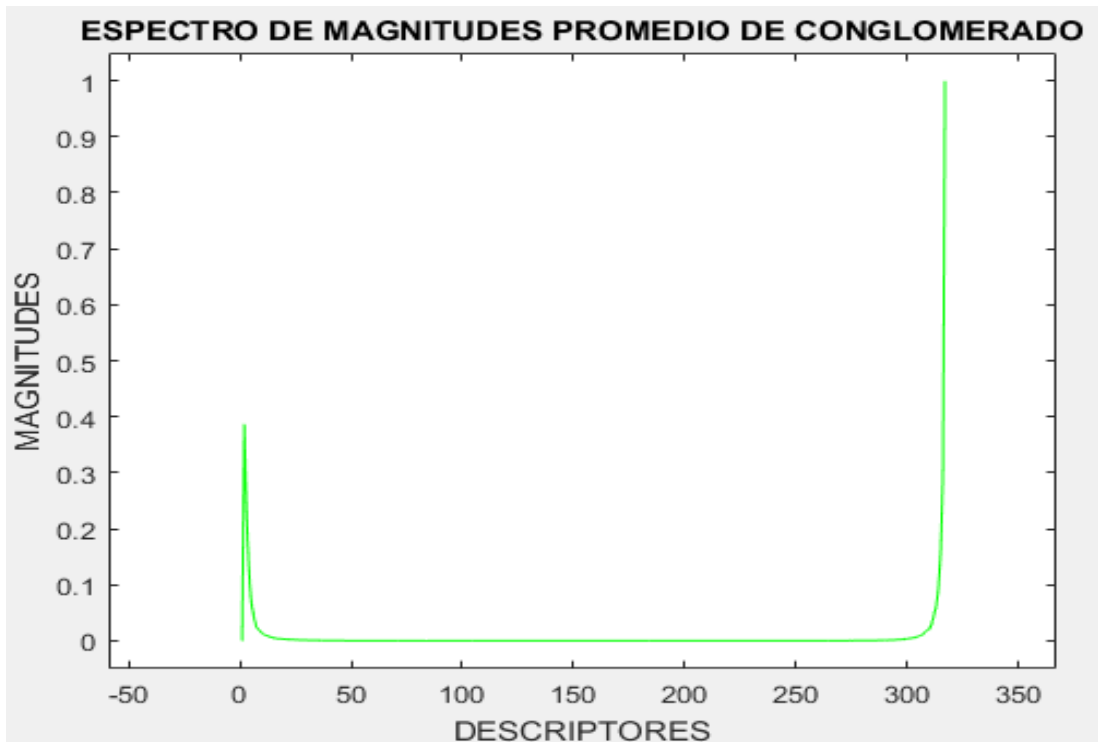


Figura 3.8 Espectro de magnitudes promedio de los 72 conglomerados.

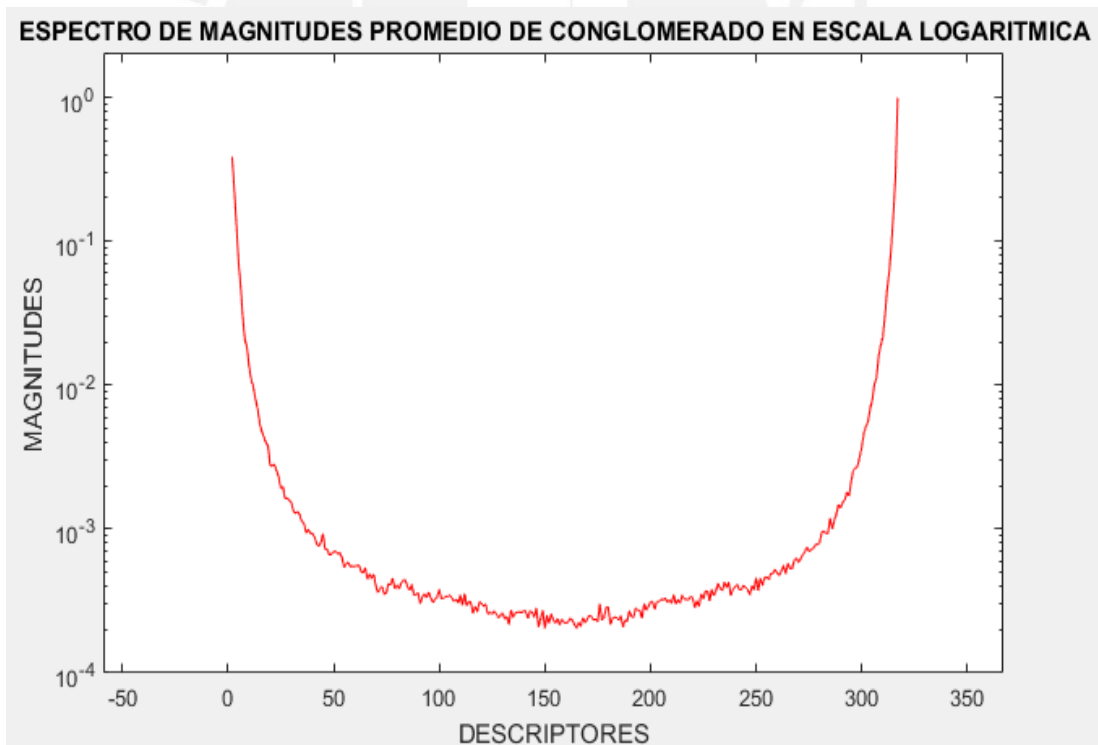


Figura 3.9 Espectro de magnitudes promedio de los 72 conglomerados en escala logarítmica



Figura 3.10 Espectro de magnitudes promedio de los 58 residuos.

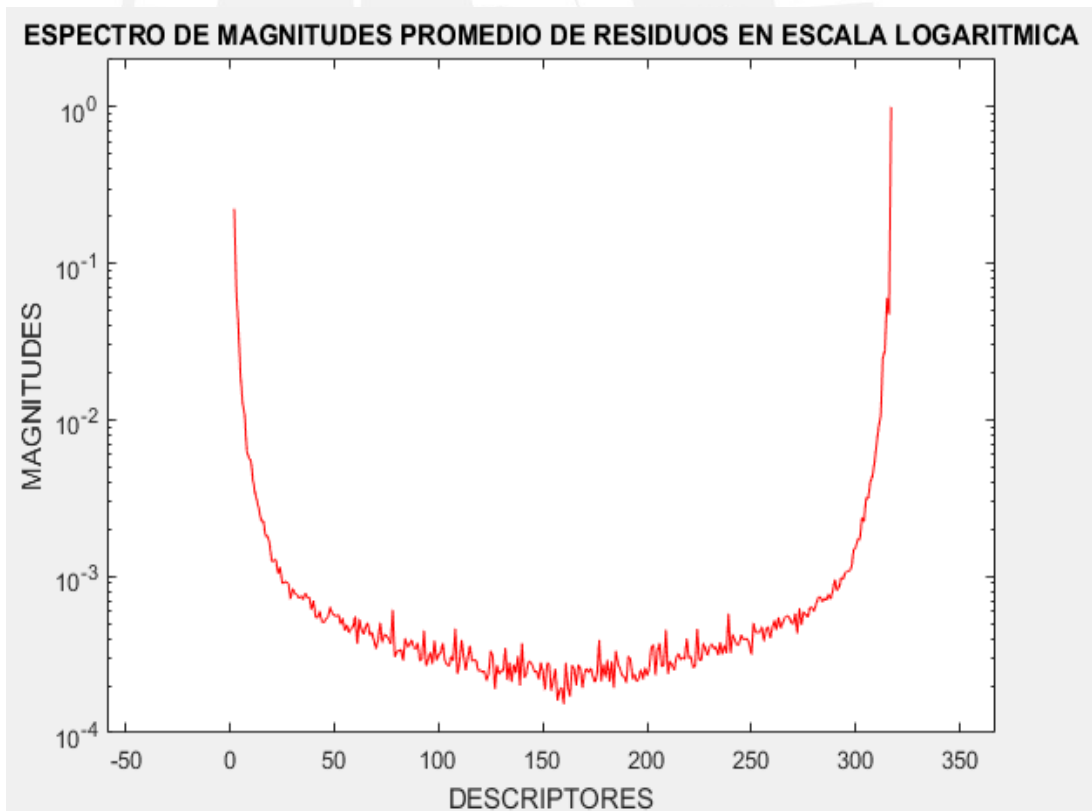
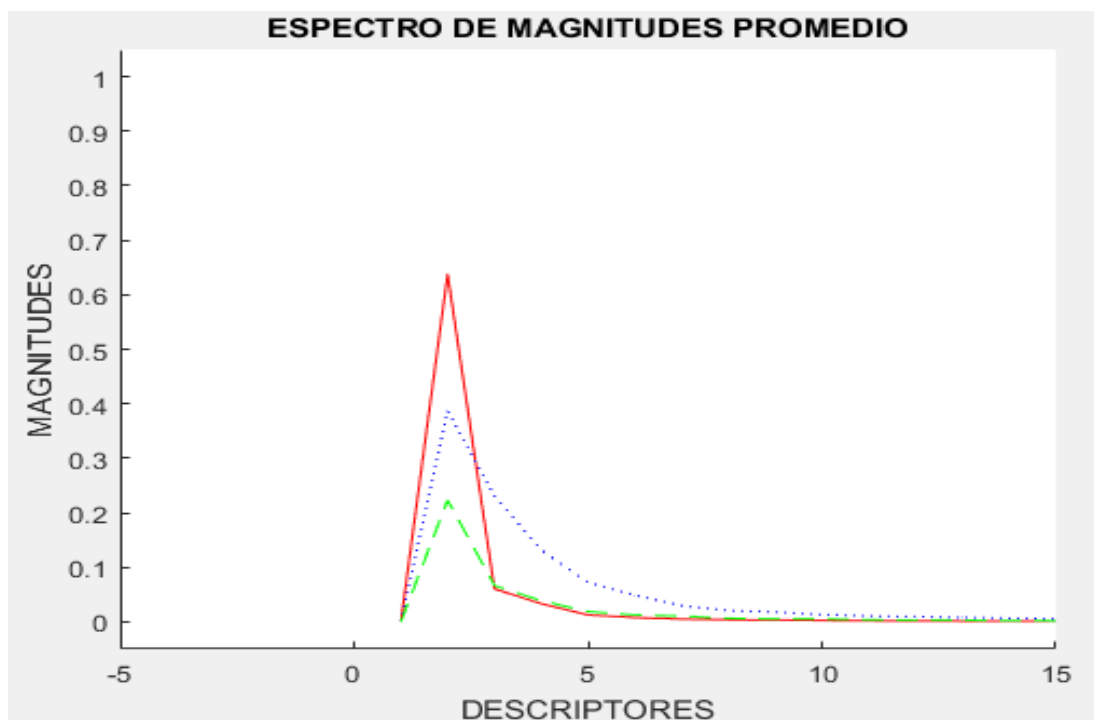
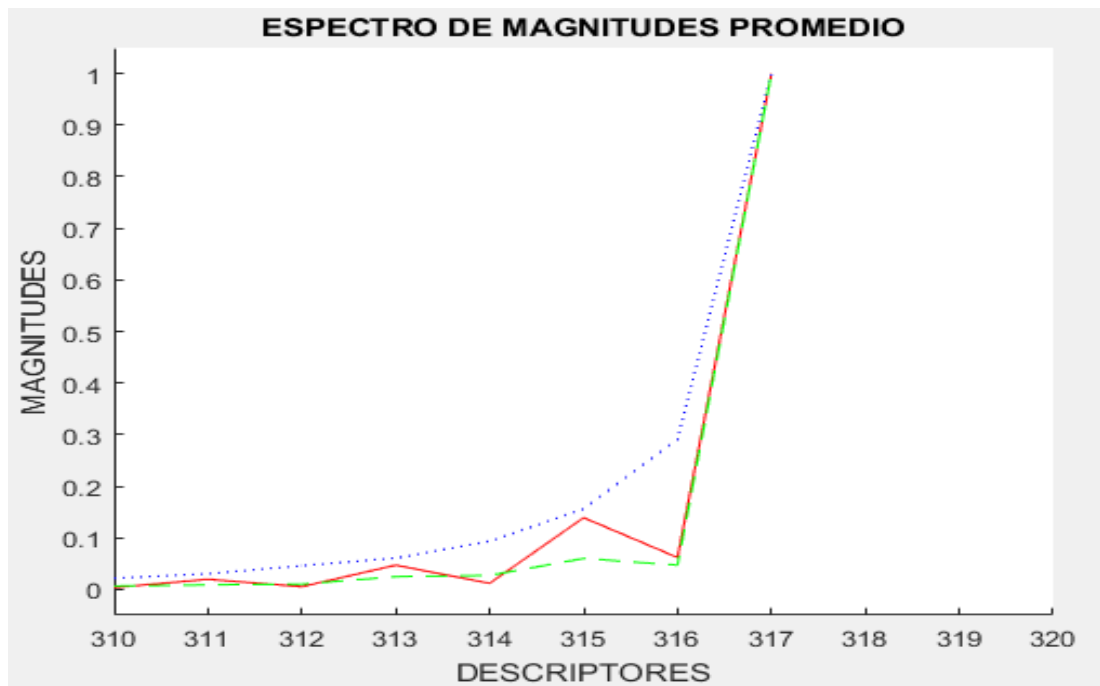


Figura 3.11 Espectro de magnitudes promedio de los 58 residuos en escala logarítmica

Se sabe que a mayor detalle mayor ancho de banda. Los espectros muestran que los conglomerados tiene mayor ancho de banda que los residuos y estos mayor que los bacilos, lo cual es congruente con la teoría. Esto se debe a las curvas pronunciadas en el caso de conglomerados y los vértices marcados en el caso de los residuos. (ver Figura 3.12)



(a)



(b)

Figura 3.12 (a)(b) Comparación de espectro de magnitudes promedio en la zona de bajas frecuencias, entre bacilos (línea continua roja), conglomerados (línea punteada azul) y residuos (línea con trazas verde).

CAPITULO 4

APLICACIÓN DE RECONOCIMIENTO DE PATRONES PARA DETECCIÓN DE BACILOS

4.1 Introducción

El presente estudio inició con una base de datos de 480 objetos, cada uno con 317 características (descriptores de Fourier). Del análisis en el capítulo anterior anotamos que las magnitudes del 1er y el último descriptor son las mismas para los tres tipos de objetos, debido a que están normalizados, por lo que no se considerarán para el estudio de selección de características, quedando finalmente 315 descriptores por objeto.

4.2 Metodología

4.2.1 Criterios de reconocimiento

Debido a que el presente estudio aplica el reconocimiento de patrones a una prueba de diagnóstico de tuberculosis, buscamos considerar los mismos criterios utilizados por el personal de salud que realiza la prueba.

Sabemos que el personal de salud, de acuerdo a la normativa, sólo contabiliza los bacilos perfectamente definidos, dejando de lado aquellos objetos que probablemente son bacilos pero al no estar perfectamente definidos pueden generar alguna duda. Por lo tanto, en el algoritmo desarrollado en el presente estudio, además del porcentaje de acierto también se ejecutó considerando la especificidad como criterio de selección con el fin de comparar los resultados. Es importante mencionar que la sensibilidad y especificidad de la baciloscopía indica la

probabilidad de clasificar correctamente a un individuo como enfermo o sano, respectivamente. Sin embargo, la sensibilidad y especificidad del algoritmo desarrollado por el autor indica la probabilidad de clasificar correctamente a una estructura como bacilo o no bacilo.

Con respecto a la cantidad de descriptores a considerar para obtener una óptima clasificación, partimos con la cantidad de descriptores de Fourier con que trabajó Veropolous [8], es decir 14 descriptores. Cabe mencionar que no son los mismos que se estudiaron en el presente estudio. Así mismo, también se ejecutó el programa buscando los 7 y 5 descriptores que mejor discriminen entre las clases.

4.2.2 Selección de las características de forma

El presente estudio busca seleccionar los descriptores que mejor discriminen entre las clases conocidas de los objetos. Como hemos apreciado, esta tarea es muy compleja debido al comportamiento de la información y la cantidad de datos a analizar. Por lo que es fundamental reducir el tamaño del vector de características a un número adecuado. El método utilizado para este fin es el método sub-óptimo de selección de características hacia atrás (BFS = Back Forward Selection).

Como fue estudiado en el Capítulo 2, el método determinístico BFS, empieza a evaluar el set completo y va extrayendo uno a uno las características en función al desempeño de un clasificador seleccionado, descartando la característica que menos aporta en cada ejercicio. En el presente estudio se optó por un clasificador basado en análisis discriminante lineal.

Para validar los resultados se aplicó la técnica de validación cruzada de tres capas. Así, se dividió la muestra total en tres sub-muestras, que contienen bacilos y no

bacilos en similar proporción que la muestra total. Una de las sub-muestras se utiliza para entrenar al clasificador seleccionado y las otras dos para validar. Este proceso se repite tres veces, con diferente sub-muestra de entrenamiento, tal como se aprecia en la Figura 4.1.

El clasificador da un resultado en base al criterio de selección (porcentaje de acierto/especificidad) por cada vez que se ejecuta. Así, se obtienen tres valores de porcentaje de acierto (y de especificidad) para la misma combinación de descriptores, el promedio de dichos valores es el valor que se le asigna al clasificador para cada combinación de descriptores. Dicho valor indicará el desempeño de cada combinación.

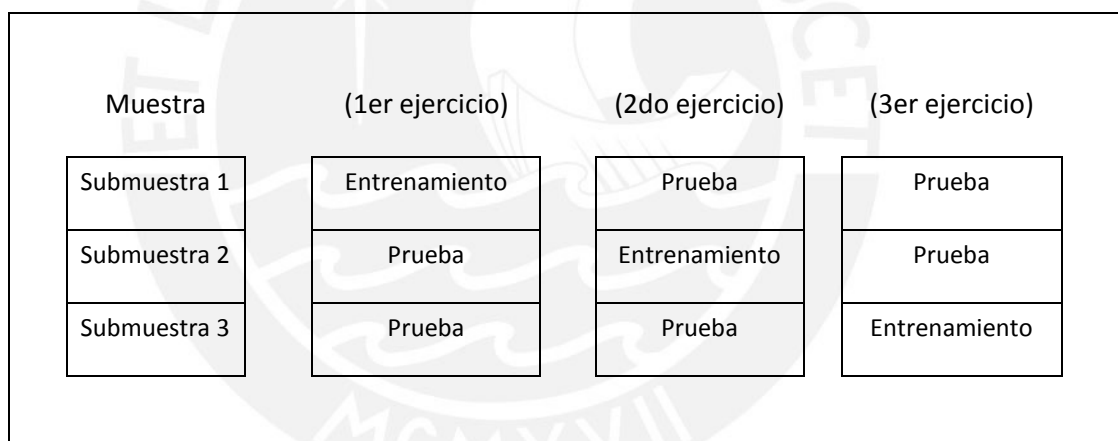


Figura 4.1 Validación cruzada de 3 capas

Algoritmo desarrollado

Se procedió a la elaboración del programa DetecciónBK, en el cual se desarrolló un algoritmo que evaluó el set completo de características de 480 x 315 y extrajo uno a uno la que menos aportaba al desempeño del clasificador (BFS = Back Forward Selection). El clasificador basado en análisis discriminante lineal que se utilizó en el programa es la función Classify, de la librería del Image Processing Toolbox, del

MATLAB, versión 7.8. El programa diseñado incluye la validación cruzada de 3 capas.

4.3 Resultados

El programa arroja como resultado los 14, 7 y 5 descriptores que mejor resultado obtuvieron en base al porcentaje de acierto y a la especificidad. Además, muestra el porcentaje de acierto, sensibilidad y especificidad de cada una de las 6 pruebas (Ver Tabla 4.1)

En la Tabla 4.1 se aprecia que los descriptores identificados bajo el criterio de acierto son casi en su totalidad diferentes de los identificados bajo el criterio de especificidad. Así mismo, mientras se considere el mismo criterio de selección, los sub-sets de menor cantidad de descriptores estarán incluidos totalmente en los sub-set de más cantidad de descriptores, lo cual tienen sentido ya que es el mismo programa que se detuvo al llegar al número de descriptores solicitado. Notamos que conforme aumenta la cantidad de descriptores, aumenta el porcentaje de acierto. Sin embargo, la relación entre dichas variables no es lineal, tal como se aprecia en la Figura 4.2.

Podemos comprobar que la combinación de descriptores de Fourier que mejor resultado arrojó, de acuerdo al algoritmo, fue la conformada por el **2, 3, 4, 9, 20, 26, 232, 259, 292, 305, 306, 313, 314 y 315** descriptor de Fourier, ya que al ejecutar el programa bajo el criterio de porcentaje de acierto se obtuvo el **96.86%** de acierto al identificar los objetos bacilos y no bacilos. Es importante mencionar que al ejecutar el programa en las tres ocasiones con el criterio de porcentaje de acierto, la sensibilidad del algoritmo siempre fue **100%** independiente de la cantidad de descriptores.

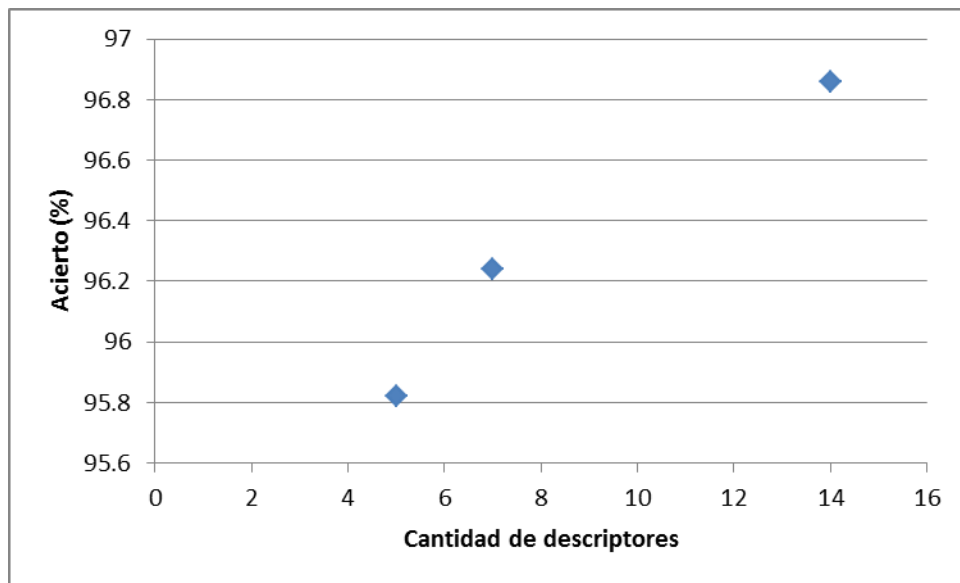


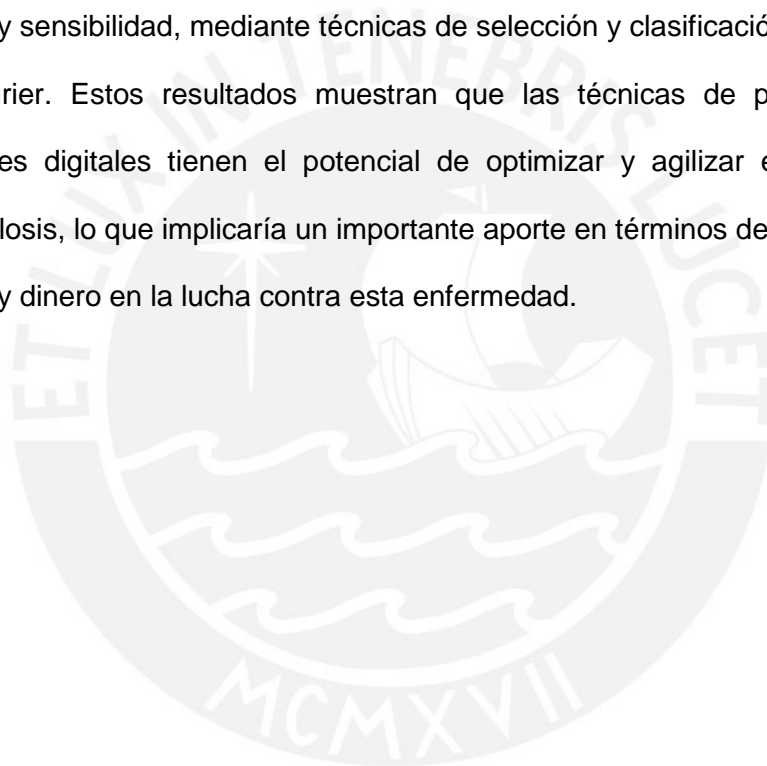
Figura 4.2 Relación entre cantidad de descriptores y porcentaje de acierto

Adicionalmente, se elaboró el programa Bacounter, el cual es capaz de recibir una cantidad de estructuras individuales segmentadas (bacilos o no bacilos) y mediante el análisis de los descriptores identificados anteriormente, es decir 14, nos indica el número de bacilos existentes en dicho grupo. El tiempo de ejecución de este programa fue de 1.7 segundos en promedio.

La mayoría de los trabajos mencionados en el Capítulo 2 mostraron sus resultados basados en la exactitud con la que sus algoritmos identifican bacilos, tal como en el presente estudio, por lo que podemos compararlos. Juan Sato[19] logra 98.79% de exactitud y un 90.1% de sensibilidad, con un algoritmo propio de segmentación mediante umbralización y posterior clasificación de bacilos en muestras de esputo teñidas bajo la técnica de Ziehl Neelsen. Así mismo, Veropolous [8] logra un 92.05% de exactitud y 84.2% de sensibilidad usando el clasificador BP en muestras teñidas con auramina y Costa [20] mediante segmentación por colores, clasificación y un filtro diseñado por los autores (ratio de color) lograron una sensibilidad de 96.8% al identificar bacilos. Por nuestro lado, se logró 96.86% de acierto y a

diferencia de los estudios previos un 100% de sensibilidad mediante descriptores de Fourier y un clasificador lineal. La ventaja de obtener una sensibilidad alta, como la obtenida, es muy importante para este tipo de estudios ya que quiere decir la cantidad de falsos negativos (bacilos no identificados como tal) es baja o nula, es decir si existe un bacilo en la muestra el sistema lo detectará.

Como resultado del presente capítulo se puede concluir que es posible la identificación de estructuras inicialmente segmentadas, con un alto porcentaje de acierto y sensibilidad, mediante técnicas de selección y clasificación de descriptores de Fourier. Estos resultados muestran que las técnicas de procesamiento de imágenes digitales tienen el potencial de optimizar y agilizar el diagnóstico de tuberculosis, lo que implicaría un importante aporte en términos de recurso humano, tiempo y dinero en la lucha contra esta enfermedad.

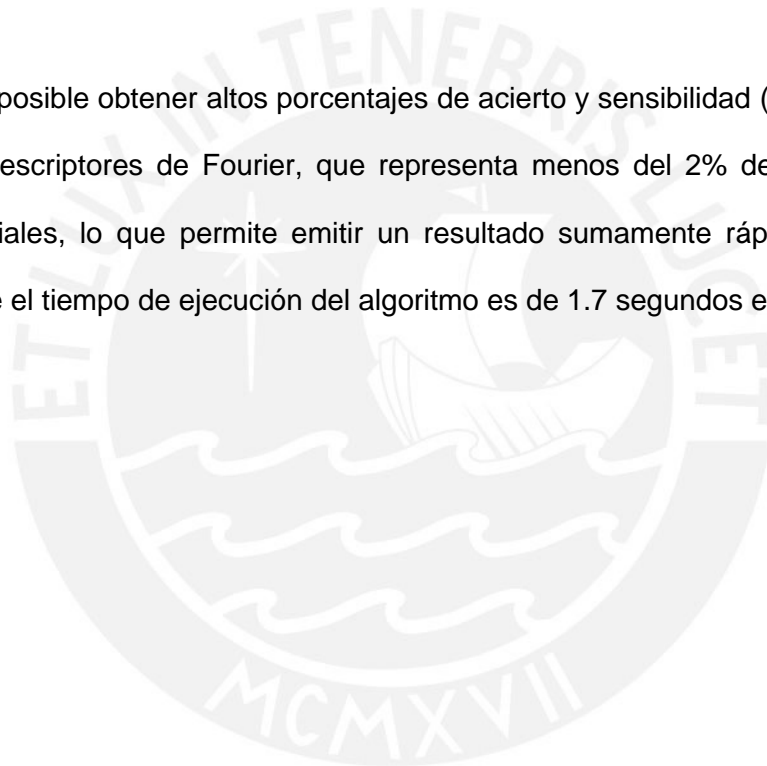


NRO DE DESC.	CRITERIO DE SELECCIÓN	DESCRIPTORES SELECCIONADOS	RESULTADOS		
			Acierto (%)	Sensibilidad (%)	Especificidad d (%)
5	Acierto	2, 4, 292, 314, 315	95.82	100	89.16
5	Especificidad	4, 10, 11, 315, 316	92.69	97.70	91.54
7	Acierto	2, 3, 4, 292, 306, 314, 315	96.24	100	89.96
7	Especificidad	4, 6, 11, 282, 306, 313, 315	93.73	98.56	94.59
14	Acierto	2, 3, 4, 9, 20, 26, 232, 259, 292, 305, 306, 313, 314, 315	96.86	100	91.47
14	Especificidad	4, 6, 8, 11, 19, 44, 154, 282, 296, 304, 306, 307, 313, 315	94.15	97.99	98.47

Tabla 4.1 Resultados al ejecutar el programa principal. Se muestran los descriptores que mejor discriminan en función del número de descriptores y el criterio de selección escogidos.

CONCLUSIONES

- Se logró el objetivo principal de desarrollar un algoritmo capaz de identificar y contar bacilos de Koch mediante selección y clasificación de descriptores de Fourier con un porcentaje de acierto de 96.86%, una sensibilidad de 100% y una especificidad de 91.47% valores útiles en la práctica ya que implica que el algoritmo identificará como bacilo a las estructuras bacilíferas, con bajo o nulo falsos negativos.
- Es posible obtener altos porcentajes de acierto y sensibilidad (>95%) a partir de 5 descriptores de Fourier, que representa menos del 2% de los descriptores iniciales, lo que permite emitir un resultado sumamente rápido considerando que el tiempo de ejecución del algoritmo es de 1.7 segundos en promedio.



IMPLICANCIA

- Actualmente la baciloscopia implica la lectura del 1% de la muestra de un paciente, y de acuerdo a la cantidad de bacilos encontrados ahí se indica mediante cruces que tan contagioso es un paciente. Por lo que al automatizar el proceso de conteo de bacilos, se podrán leer todos los campos de la muestra, siendo necesario establecer nuevos rangos que definirán si el paciente es de 1,2 o 3 cruces.

RECOMENDACIONES

- Realizar un estudio con imágenes de esputo donde no se haya usado el azul de metileno en el teñido de la muestra, con el fin de comparar los resultados. Si los resultados son similares o mejores significaría un importante ahorro al estado en el insumo azul de metileno.
- Realizar un estudio con una base de datos de mínimo 100 pacientes con diagnóstico validado por la prueba de cultivo (gold standard), lo que permitirá comparar la sensibilidad y especificidad del programa con la baciloscopia.
- Analizar otros descriptores de forma para optimizar la descripción y la clasificación, reduciendo la información en cantidad más no en calidad.
- Ejecutar el algoritmo Bacounter con estructuras segmentadas por software y previamente estandarizadas a 317 descriptores normalizados.

BIBLIOGRAFIA

- [1] MANDELL, Gerald et al.
1991 Enfermedades infecciosas principios y practica. 3ª Edición. Buenos Aires: Editorial Médica Panamericana. Tomo 1.
- [2] VIRELLA, Gabriel
1997 Microbiology and infectious diseases. 3ª Edición. Maryland (USA): Port City Press.
- [3] WHO
2015 Global tuberculosis Report 2015 WHO Report 2015. Geneva-Switzerland. World Health Organization
- [4] MINISTERIO DE SALUD
2015 Situación de la tuberculosis en el Perú
- [5] JOSE A. CAMINERO LUNA
2003 Guía de la tuberculosis para Médicos Especialistas
- [6] LALVANI, Ajit
2007 Diagnosing tuberculosis infection in the 21st century: new tools to tackle an old enemy
- [7] MINISTERIO DE SALUD, INSTITUTO NACIONAL DE SALUD, CENTRO NACIONAL DE LABORATORIOS EN SALUD PÚBLICA.
1995 Manual de normas y procedimientos de la baciloscopia para el diagnóstico bacteriológico de la tuberculosis. Norma técnica nº10.
- [8] VEROPOULOS, Konstantinos
2001 Machine learning approaches to medical decision making. Tesis (PhD). University of Bristol. Facultad de Ingeniería. Bristol (Inglaterra) 265p.
- [9] FORERO, M.
2006 Automatic identification of Mycobacterium tuberculosis by Gaussian mixture models. University of Birmingham. Escuela de Biociencias. Birmingham (Inglaterra).
- [10] THEODORIDIS, Sergios;
2009 Pattern Recognition. 4ª Edición, San Diego, California. Academic Press.
- [11] GONZÁLEZ, Rafael; Woods, Richard; Eddins, Steven
2004 Digital Image Processing using Matlab. New Jersey. Pearson Prentice Hall.
- [12] GONZÁLEZ, Rafael ; Woods, Richard
1996 Tratamiento digital de imágenes. Wilmington, Delaware, USA. Addison-Wesley Iberoamericana S.A.
- [13] SCHLOSSBERG
1999 Tuberculosis and nontuberculous mycobacterial infections. 4th Edition. Philadelphia: Saunders Company.

- [14] ALEXANDROV ,V.V; Gorsky, N.D
1991 From humans to computers - Cognition through visual perception.
Singapore. World Scientific Publishing
- [15] HAITAO, Jia;
2005 "Improvement of Fourier Descriptors using spatial normalization",
Conference: IEE International Symposium on Communications and
information Technology (ISCIT 2005). Volume 2, pp. 1284-1287
- [16] REEVES, Anthony; PROKOP, R.J; ANDREWS, Susan; KUHL, Frank;
1998 Three-Dimensional Shape Analysis using moments and Fourier
descriptors. IEE Transactions on Pattern Analysis and Machine
Intelligence, Volume 10, pp 937-943
- [17] SADAPHAL, P., RAO, J., COMSTOCK, G. W. BEG, M. F. ;
2008 "Image processing techniques for identifying Mycobacterium
tuberculosis in Ziehl-Neelsen stains". International Journal of
Tuberculosis Lung Disease. Paris, vol. 12, n. 5, pp. 579-582
- [18] SOTAQUIRÁ, M., RUEDA, L., NARVAEZ, R. ;
2009 "Detection and quantification of bacilli and clusters present in sputum
smear samples: a novel algorithm for pulmonary tuberculosis
diagnosis". Proceedings of international Conference on Digital Image
Processing. Bangkok, Tailandia, pp. 117-121
- [19] SATO, JUAN;
2013 Microscopio automatizado: conteo de bacilos de tuberculosis. Tesis
para optar el Título de ingeniero Electrónico. Lima: Pontificia
Universidad Católica del Perú, Facultad de Ciencias e Ingeniería.
- [20] COSTA FILHO, C., LEVY, P., XAVIER, C., FUJIMOTO, L., COSTA, M. ;
2015 "Automatic identification of tuberculosis mycobacterium". Research on
Biomedical Engineering vol. 31 n. 1, pp.33-43.