



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**

**SOFT-WARMI: EVALUACIÓN DEL  
SOFTWARE AUTOMÁTICO DE  
DIAGNÓSTICO PARA VAGINOSIS  
BACTERIANA (VB)**

**TESIS PARA OPTAR EL GRADO DE  
MAESTRO EN INFORMÁTICA BIOMÉDICA  
EN SALUD GLOBAL CON MENCIÓN EN  
INFORMÁTICA EN SALUD**

**ALICIA KATHERINE ALVA MANTARI**

**LIMA - PERÚ**

**2022**



**Asesor**

Dr. Mirko Zimic, PhD. MSc. MHS.

**Co-Asesor**

Dr. Cesar Carcamo, MD. MPH. PhD.

**JURADO DE TESIS**

DR. Segundo Cecilio Acho Mego

PRESIDENTE

MG. Daniel Flavio Condor Camara

VOCAL

MG. Pablo Alejandro Fonseca Arroyo

SECRETARIO (A)

*A mis hijas Camila y Sami motor de mi vida,  
razón de mi esfuerzo y perseverancia.*

*A mi querido esposo Jorge Farfan por su apoyo siempre.*

*Pero sobre todo a mi madre Vilma Mantari de Alva por su lucha  
constante por defender conmigo mi derecho a estudiar y a superarme.*

*A mi papá Oswaldo Alva Yataco que su fuerza siempre me acompaña  
y abraza en momentos de debilidad.*

*A mi asesor Dr. Mirko Zimic por confiar en mí, por apoyarme y  
brindarme la oportunidad de hacer investigación y su asesoría en la  
realización del presente estudio.*

*A la Dra. Patricia García y al Dr. César Cárcamo por darme la  
esperanza de poder seguir haciendo investigación con el desafío que  
implicaba ser madre y por ser mi inspiración para seguir trabajando  
e investigando en salud pública.*

## **AGRADECIMIENTOS**

Agradezco a Dios y mi familia por su amor y apoyo constante a lo largo de esta etapa y cristalización de un sueño, en especial a Camila Farfan y Sami Farfan motivos de mi lucha, perseverancia y felicidad. A mi esposo Jorge Farfan por el apoyo constante para no rendirme a pesar de las dificultades.

Agradezco a mi madre Vilma Mantari de Alva por entender mis ansias de superarme, de estudiar y por luchar conmigo hasta lograr mi objetivo, por creer en mis esfuerzos y respaldarlos a pesar de las críticas. A mi padre Oswaldo Alva Yataco por el apoyo y la fuerza siempre desde el cielo.

Agradezco a mi asesor, el Dr. Mirko Zimic Peralta, por darme la oportunidad de investigar aquella primera vez, mostrándome y enseñándome en cada paso del camino a lograr objetivos y no dejarnos vencer. Por su apoyo y asesoramiento en el estudio y realización de este trabajo, por ser una guía constante y un modelo a seguir.

A los Doctores Dra. Patricia García y Dr. Cesar Cárcamo por brindarme la oportunidad de seguir soñando y haciendo investigación aun con el desafío de ser madre. Por ser mi inspiración en la investigación en Salud Pública apoyando a las poblaciones más vulnerables.

Finalmente, agradezco al QUIPU - Centro Andino de Investigación y Entrenamiento en Informática para la Salud Global de la Universidad Peruana Cayetano Heredia, al director del programa y al equipo de coordinación de la maestría por darme la oportunidad de formar parte de este programa.

## **FUENTES DE FINANCIAMIENTO**

Esta investigación fue financiada por QUIPU - Centro Andino de Investigación y Entrenamiento en Informática para la Salud Global de la Universidad Peruana Cayetano Heredia

# SOFT-WARMI: EVALUACIÓN DEL SOFTWARE AUTOMÁTICO DE DIAGNÓSTICO PARA VAGINOSIS BACTERIANA (VB)

## INFORME DE ORIGINALIDAD

<b>8%</b>	<b>8%</b>	<b>2%</b>	<b>2%</b>
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

## FUENTES PRIMARIAS

<b>1</b>	<b>www.scielo.org.pe</b> Fuente de Internet	<b>1%</b>
<b>2</b>	<b>repositorio.ug.edu.ec</b> Fuente de Internet	<b>1%</b>
<b>3</b>	<b>hdl.handle.net</b> Fuente de Internet	<b>1%</b>
<b>4</b>	<b>www.salud.pr.gov</b> Fuente de Internet	<b>&lt;1%</b>
<b>5</b>	<b>www.scielo.br</b> Fuente de Internet	<b>&lt;1%</b>
<b>6</b>	<b>repositorio.unap.edu.pe</b> Fuente de Internet	<b>&lt;1%</b>
<b>7</b>	<b>Submitted to Universidad Católica de Santa María</b> Trabajo del estudiante	<b>&lt;1%</b>
<b>8</b>	<b>doku.pub</b> Fuente de Internet	<b>&lt;1%</b>

## Tabla de Contenidos

RESUMEN

ABSTRACT

1. INTRODUCCIÓN .....	1
2. PLANTEAMIENTO DE LA INVESTIGACIÓN .....	3
2.1. Planteamiento del problema. ....	3
2.2. Justificación del estudio .....	4
2.3. Objetivos del estudio .....	5
3. MARCO TEÓRICO .....	6
3.1. Patogénesis de la Vaginosis Bacteriana (VB) .....	6
3.2. Epidemiología: estadísticas en el mundo y en el Perú .....	9
3.3. Situación de la VB en el Perú problemática de la VB en la salud pública ...	10
3.4. Tratamiento de la VB diagnóstico de la VB .....	11
3.5. Reconocimiento de patrones y sus aplicaciones en diagnóstico de enfermedades telemedicina y telediagnóstico .....	12
4. HIPÓTESIS .....	14
5. METODOLOGÍA .....	15
5.1. Diseño de estudio .....	15
5.2. Población y muestra .....	15
5.3. Procedimientos para desarrollo de SoftWarmi .....	15
5.3.1.Obtención de fotografías .....	18
5.3.2.Procesamiento de imagen .....	20
5.3.3.Extracción de características .....	26
5.3.4.Análisis de clasificación .....	35
5.3.4.1. Modelo de objetos – Nivel 1 .....	36
5.3.4.2. Modelo de fotos – Nivel 2 .....	38

5.3.4.3. Modelo de muestras – Nivel 3 .....	39
5.3.5.Consideraciones éticas .....	40
5.3.6.Recursos Requeridos .....	40
6. RESULTADOS .....	42
6.1. Procesamiento de imagen .....	42
6.2. Reconocimiento de patrones .....	43
6.2.1.Modelo objetos – Nivel 1 .....	43
6.2.1.1. Modelo <i>Lactobacillus</i> .....	44
6.2.1.2. Modelo <i>Gardnerella</i> .....	47
6.2.1.3. Modelo <i>Mobiluncus</i> .....	50
6.2.1.4. Resumen de Modelos .....	53
6.2.1.5. Evaluación secuencial de modelos de objeto .....	56
6.2.2.Modelo de fotos - Nivel 2 .....	59
6.2.3.Modelo de muestra - Nivel 3 .....	66
7. DISCUSIÓN .....	69
8. LIMITACIONES .....	75
9. CONCLUSIONES Y RECOMENDACIONES .....	76
10. BIBLIOGRAFÍA .....	78

## **RESUMEN – PALABRAS CLAVES**

La vaginosis bacteriana (VB) puede considerarse un problema de salud pública debido a la cantidad de mujeres afectadas por ella. Esto es debido a que es asintomática, en la mayoría de casos. Los problemas asociados a ella corresponden principalmente a la vulnerabilidad a la salud femenina al contraerla. La vaginosis bacteriana es alarmante principalmente durante los primeros tres meses de gestación, pues hace latente un riesgo de ruptura de membranas o abortos espontáneos entre otros. Cabe mencionar que las mujeres sexualmente activas además aumentan sus posibilidades de contraer y transmitir enfermedades de transmisión sexual (ETS). Uno de los principales problemas asociados a VB, tiene que ver con el diagnóstico mismo de la infección. El diagnóstico es altamente variable entre el personal entrenado para esta labor, razón por la cual es importante contar con herramientas que permitan un diagnóstico más rápido y confiable y menos variable referente a VB. Este estudio desarrollo una herramienta computacional para realizar tal diagnóstico, y además implemento un medio de comunicación masiva para hacer acceder remotamente a él.

SoftWarmi es un software dedicado al diagnóstico de VB, de manera remota a partir de 10 fotos de la muestra. Este algoritmo se basa en el conteo de los morfotipos en la muestra de manera similar a lo realizado por el score de Nugent, el método utilizado de manera tradicional para el diagnóstico de VB.

Se analizaron 59 láminas de hisopados vaginales. Las láminas estaban clasificadas por 30 positivas y 29 negativas a VB. De cada muestra han sido capturados 10 campos, seleccionadas por el Blga. Lorena Rojas, entrenada para tal fin

anteriormente. De cada fotografía se han extraído un total de 500 objetos aprox. De cada objeto se han calculado un conjunto de 184 características cuantificables de forma, color, textura e iluminación. Posteriormente se diseñaron modelos estadísticos para el reconocimiento de Lactobacillus, Gardnerellas y Mobiluncus. Para Lactobacillus obtuvimos una sensibilidad de 97.63% y una especificidad de 99.10%. Para el reconocimiento de Gardnerella se alcanzó una sensibilidad de 93.64% y una especificidad de 97.40%. Por ultimo para Mobiluncus alcanzamos una sensibilidad de 96.23% y una especificidad de 95.51%. Luego se realizaron conteos de cada morfotipo de manera secuencial y excluyente, evaluándolas según la incidencia de existencia en la población general. Luego se calculó un modelo estadístico de diagnóstico por fotos, el cual alcanzo una sensibilidad de 100% y una especificidad de 96.15%, en el training set (50 fotos), y el en testing set en un conjunto de 66 fotos una sensibilidad de 94.44% y una especificidad de 100%. Al final se evaluaron 10 fotos por muestra, de 59 muestras estudiadas se alcanzó para este modelo una sensibilidad de 96.67% y una especificidad de 100%, con un porcentaje de agreement con los expertos de la universidad de Washington de 98.31% y un índice Kappa de 0.9661.

Esperamos que este sea el inicio de una herramienta para lograr que la VB, sea analizada, diagnosticada y tratada en diferentes poblaciones que no poseen los medios para poder acceder sobre todo en las zonas más alejadas, pobres y remotas del país.

Palabras claves: Reconocimiento de patrones, vaginosis bacteriana, diagnostico, software.

## **ABSTRACT**

Bacterial vaginosis (BV) can be considered a public health problem due to the number of women affected by it. This is because it is asymptomatic, in most cases. The problems associated with it correspond mainly to the vulnerability of women's health to contracting it. Bacterial vaginosis is alarming mainly during the first three months of pregnancy, as it poses a latent risk of rupture of membranes or miscarriages, among others. It is worth mentioning that sexually active women also increase their chances of contracting and transmitting sexually transmitted diseases (STDs). One of the main problems associated with BV is the diagnosis of the infection itself. Diagnosis is highly variable among trained personnel, which is why it is important to have tools that allow a faster, more reliable and less variable diagnosis of BV. This study developed a computational tool to perform such diagnosis, and also implemented a means of mass communication to make it remotely accessible.

SoftWarmi is a software dedicated to the diagnosis of BV, remotely from 10 photos of the sample. This algorithm is based on counting morphotypes in the sample in a similar way to Nugent's score, the traditional method used for BV diagnosis.

Fifty-nine slides of vaginal swabs were analyzed. The slides were classified as 30 positive and 29 negatives for BV. From each sample, 10 fields were captured and selected by Blga. Lorena Rojas, trained for this purpose previously. A total of approximately 500 objects were extracted from each photograph. A set of 184 quantifiable features of shape, color, texture and illumination were calculated from each object. Subsequently, statistical models were designed for the recognition of

Lactobacillus, Gardnerella and Mobiluncus. For Lactobacillus we obtained a sensitivity of 97.63% and a specificity of 99.10%. For the recognition of Gardnerella, we reached a sensitivity of 93.64% and a specificity of 97.40%. Finally, for Mobiluncus we achieved a sensitivity of 96.23% and a specificity of 95.51%. Counts of each morphotype were then carried out sequentially and in an exclusionary manner, evaluating them according to the incidence of occurrence in the general population. Then a statistical model of diagnosis by photos was calculated, which reached a sensitivity of 100% and a specificity of 96.15%, in the training set (50 photos), and in the testing set in a set of 66 photos a sensitivity of 94.44% and a specificity of 100%. In the end, 10 photos per sample were evaluated, and out of 59 samples studied, a sensitivity of 96.67% and a specificity of 100% were achieved for this model, with a percentage of agreement with the experts of the University of Washington of 98.31% and a Kappa index of 0.9661.

We hope that this is the beginning of a tool to ensure that BV is analyzed, diagnosed and treated in different populations that do not have the means to access it, especially in the most remote, poor and remote areas of the country.

**Keywords:** Pattern recognition, bacterial vaginosis, diagnosis, software.

## 1. INTRODUCCIÓN

La vaginosis bacteriana (VB) es una infección muy frecuente entre las mujeres sexualmente activas y genera altos costos en atención médica. La VB es la principal causa de infección vaginal en mujeres en edad reproductiva, aproximadamente el 50% de las pacientes pueden cursar asintomáticas(1). La VB se convierte en un problema de salud global por el gran número de infectadas cada año y por las complicaciones que presenta sobre todo en madres gestantes(2) (abortos(3), rotura de membranas(4), embarazos ectópicos (5), parto prematuro(6), etc.) además la vulnerabilidad a la que expone a la mujer a enfermedades de transmisión sexual entre las cuales destacan el VIH, y las ETS en general(7). La VB es una alteración de la flora vaginal bacteriana normal. En el caso de VB, se identifica una disminución de la concentración de *Lactobacillus* (“bacterias beneficiosas”) y un incremento de bacterias patógenas (“bacterias dañinas”) como son *Gardnerella*, *Mobiluncus*, *Bacteroides*, *Prevotellas*, entre otros(8). El diagnóstico de VB se calcula mediante el **Score de Nugent**, que requiere la observación microscópica de láminas coloreadas con tinción de Gram. El criterio de Nugent(9) se basa en una escala ordinal, correspondiendo a resultados desde el puntaje de 0 hasta el puntaje de 10. **Una desventaja de esta metodología es la subjetividad de las lecturas realizadas por el personal técnico de laboratorio, y el tiempo de entrenamiento requerido para lograr las habilidades para una lectura adecuada.**

En nuestro país la mayor parte de la ciudadanía no cuenta con recursos médicos ni económicos para realizarse un examen, además en muchos sectores pobres y remotos no existe un experto que realice el examen, es importante recalcar que los

especialistas no tiene el debido entrenamiento en las lecturas de VB, y además existe una alta variabilidad entre lectores.

Dado el estado crítico del diagnóstico de VB es necesario contar con alguna herramienta que permita revertir esta situación. Es por esto la importancia de contar con un sistema de telemedicina que permita el diagnóstico remoto de VB. Se propone evaluar un software de diagnóstico para VB basado en el reconocimiento de patrones geométricos, de iluminación y color para los morfotipos presentes en las muestras proponiéndola en un futuro como herramienta que brinde la confiabilidad y elimine la subjetividad del diagnóstico humano.

El proyecto tiene como objetivo resolver la principal limitación para facilitar el diagnóstico de VB de manera automática.

Este proyecto tiene como objetivo optimizar y evaluar un software automático SoftWarmi para el diagnóstico de vaginosis bacteriana, que implementado de manera automática en una aplicación web podría ser usado en lugares remotos de escasos recursos.

## **2. PLANTEAMIENTO DE LA INVESTIGACIÓN**

### **2.1. Planteamiento del problema.**

#### *Vaginosis un problema de salud global y Nacional*

La VB es la infección vaginal más frecuente en el mundo. Los valores de prevalencia varían entre los valores de 5 a 51%, el valor depende de las características demográficas y culturales (10). En nuestro país Sánchez y col. (11) dieron cifras de 30%, lo que ha sido comprobado, en estudios locales con menos población(10).

#### *Complicaciones de VB*

La VB expone a las mujeres a graves riesgos asociadas a ella como son: mayor susceptibilidad a contraer HIV , mayor posibilidad de transmitir HIV (7); aumenta la susceptibilidad de la mujer a contraer otras ETS(5).

La VB, genera complicaciones durante la gestación principalmente en los primeros tres meses de gestación(6) (abortos, ruptura prematura de membranas, embarazos ectópicos, parto prematuro, etc.). Esto es preocupante debido a que un informe presentado por el Hospital nacional Cayetano Heredia (UPCH) en 2002 (12) registro que más de 50% de la vaginosis bacteriana en gestantes es asintomática, razón por la cual no es diagnosticada ni recibe tratamiento(1). Por la importancia y la trascendencia de las complicaciones de una infección por VB se concluye la importancia de contar con un diagnóstico efectivo y al alcance de la mayoría de centros de salud a nivel nacional.

Los métodos de diagnóstico actuales combinan verificación clínica y microbiológica. Entre las pruebas microbiológicas, están el método de Nugent basado en la observación microscópica directa.

Entre algunos de los problemas con el diagnóstico tenemos:

- Ausencia de Personal técnico de salud que diagnostiquen correctamente
- Especialistas no lo consideran importante (sensibilidad al problema)
- Falta de recursos (láminas, guantes etc.).
- No hay entrenamiento adecuado.
- No saben realizar la toma de muestras en el lugar de atención.
- Si se hace exámenes no son diagnosticados pues no hay personal capacitado para este fin.

En un estudio preliminar que hemos realizado comparamos las lecturas de un profesional de Lima y la Especialista Dr. Karol Spiegel(13), contando con un porcentaje de coincidencia de 45%, durante el desarrollo del proyecto “PREVEN” se ha evaluado coincidencias en lecturas de frotis vaginales antes, obteniendo resultados similares, por lo tanto tenemos referencia de la alta variabilidad de las lecturas en el diagnóstico de VB.

## **2.2. Justificación del estudio.**

La VB no es problema con alta morbilidad pero si es un problema con una alta prevalencia, asociada a muchas enfermedades especialmente de transmisión sexual(14). Dada la importancia y la trascendencia de cada una de las complicaciones ocasionadas por una infección de VB se concluye la importancia de contar con un diagnóstico efectivo(1). Es uno de nuestras visiones brindar al alcance de la mayoría de establecimientos de salud a nivel nacional este diagnóstico automático.

Debido a que no existen medios suficientes ni médicos patólogos para el diagnóstico en los lugares remotos del país y a la carencia de algún método

cuantitativo y objetivo para la lectura e interpretación de las láminas con la tinción de los frotis vaginales este diagnóstico se encarece y fácilmente es no tomado en cuenta lo que se manifiesta en riesgos en el binomio madre niño durante la gestación y problemas relacionados con la mujer en general.

Este proyecto propone salvar estos inconvenientes mediante un software automático, efectivo, gratuito y de fácil uso, con un diagnóstico de VB con una alta sensibilidad y especificidad. Con esto se obvia el especialista in situ necesario para hacer el análisis, de tal modo que cada centro de salud en el país solo requiere un microscopio directo, los componentes usados en la tinción de Gram bastante accesibles, y una cámara con una computadora lo cual le permitirá contar con un sistema de diagnóstico. Se busca por eso optimizar y evaluar el software de diagnóstico automático SoftWarmi que podría ser utilizado desde cualquier lugar del mundo y así obtener un diagnóstico de VB confiable seguro y rápido.

### **2.3. Objetivos del estudio.**

**Objetivo General:** Optimizar y evaluar el software automático de diagnóstico de vaginosis bacteriana **SoftWarmi**, un software basado en el reconocimiento de morfotipos presentes en una infección por VB.

### 3. MARCO TEÓRICO

#### 3.1. Patogénesis de la Vaginosis Bacteriana (VB).

La vaginosis bacteriana (VB) es la infección vaginal más frecuente en las mujeres en edad de fértil y de alta frecuencia en mujeres embarazadas((15),(16),(3)). En las mujeres embarazadas el equilibrio bacteriano normal se ve alterado y en su lugar ciertas bacterias crecen de manera excesiva, lo que hace a las gestantes más a proclives a esta enfermedad. Los síntomas pueden ser variados, presencia de flujo vaginal anormal con olor fétido característico a "pescado" ((17), (16)), color blanco o gris y puede ser poco espeso. Ardor al orinar o picazón en la parte externa de la vagina o ambos síntomas, sin embargo la mayoría de las mujeres indican no tener síntomas(16).

La causa de la vaginosis bacteriana no es clara. Normalmente, la mayoría de las bacterias en la vagina son “buenas”, entre ellas las más frecuentes son los *Lactobacillus* (Figura1.[(a)]). En ocasiones, en la vagina se pueden encontrar ciertas bacterias que son “dañinas” (Figura1.[(e)-(f)]). La vaginosis bacteriana no se puede considerar una vaginitis, es una alteración de la flora vaginal bacteriana normal donde hay una reducción de las bacterias beneficiosas los *Lactobacillus* que son productores de peróxido de hidrógeno y un incremento de prevalencia y concentración de las bacterias no beneficiosas como *Gardnerella Vaginalis*(45-99%), *Mycoplasma hominis*(M. hominis (58- 76%), y los anaerobios: *Mobiluncus* Sp(40- 60%), *Prevotella* Sp, *Bacteroides* Sp. y *Peptoestreptococo* ( (18) ,(16),(19),(20),(21)) (62- 92%).

La vaginosis bacteriana se presenta cuando hay un aumento del número de bacterias dañinas y una disminución de las beneficiosas. Se desconoce los métodos de transmisión o condiciones para contraer la vaginosis bacteriana. Aún así, si se conocen algunas actividades alteran el equilibrio normal de las bacterias vaginales y aumentan el riesgo de exponer a la mujer de contraer vaginosis bacteriana, entre ellas podemos citar: Poseer una nueva pareja sexual o múltiples parejas sexuales, utilizar duchas vaginales, dispositivos intrauterinos, tabaquismo, o un bajo nivel socioeconómico.

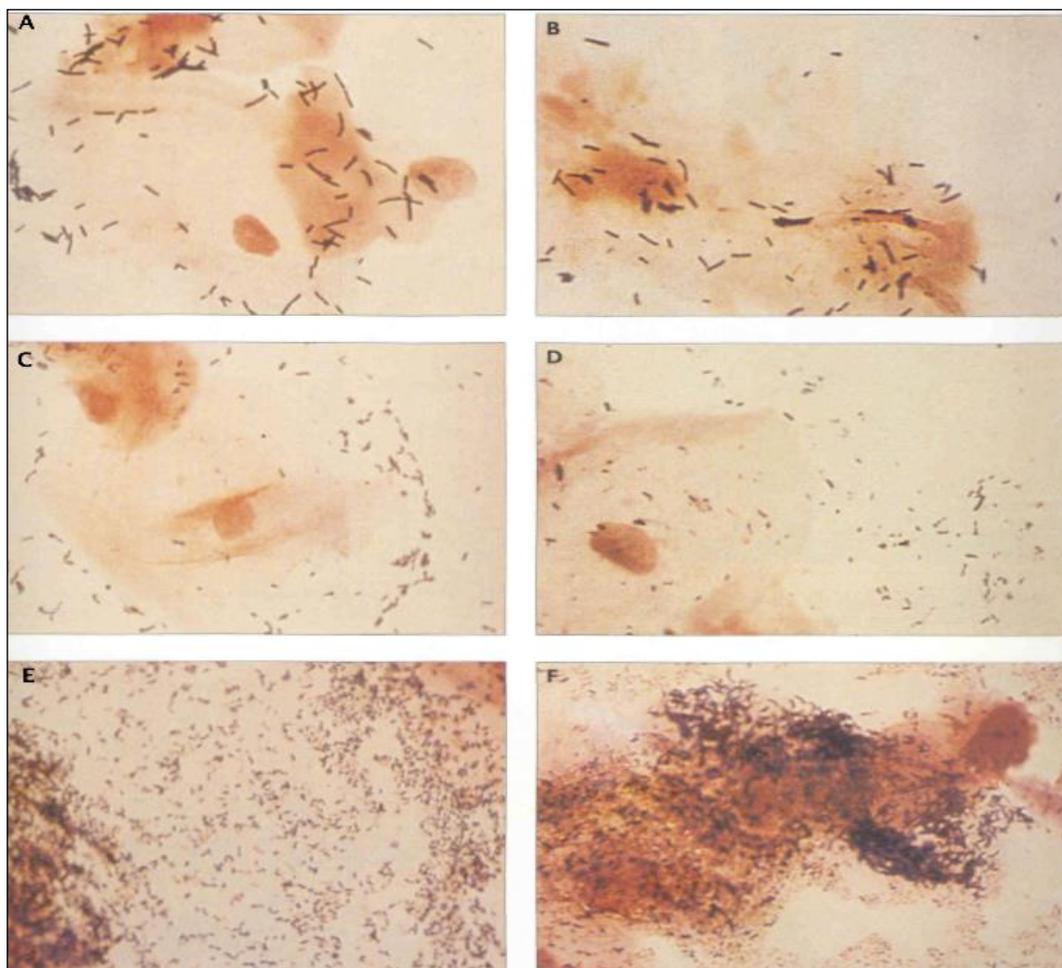
Hasta un 50% de las pacientes pueden cursar asintomáticas (22),(16) .Es importante diagnosticar y tratar oportunamente las infecciones vaginales ellas generar complicaciones graves; la VB está asociada con el parto pretérmino (23), ruptura prematura de membranas(4), corioamnionitis(24),(25), endometritis puerperal (26), enfermedad inflamatoria pélvica(27),(28) e infecciones postoperatorias (29),infertilidad(30), y cáncer cervical (14) entre otras.

### ***Diagnóstico Vaginosis Bacteriana***

El diagnóstico se basa en el Criterio de Amsel (17), donde se la presencia de tres de los cuatro criterios siguientes: 1) Secreción vaginal homogénea, 2) PH vaginal superior a 4.5, 3) Un "olor" característico (olor a pescado) debido a la liberación de aminas ante la adición de una base, y 4) presencia de numerosas bacterias fijadas sobre la superficie de las células epiteliales vaginales que llegan a oscurecer su borde constituyen lo que se ha llamado células guía. La tinción Gram se considera por muchos como el “estándar de oro” para el diagnóstico de vaginosis bacteriana, sin embargo su interpretación requiere

experiencia, y con frecuencia es difícil obtener resultados en el momento oportuno para el diagnóstico clínico (Figura1). El diagnóstico clínico de la vaginosis bacteriana usando el criterio de Nugent (2) está dado en el cuadro (Figura2) este ofrece un resultado en una escala ordinal, correspondiendo a resultados desde el puntaje de 0 hasta el puntaje de 10.

En la Figura 1 observamos un frotis vaginal con Tinción de Gram siguiendo el diagnóstico por el criterio de Nugent podemos observar el diagnóstico para cada campo en la figura



**Figura 1. Imágenes microscópicas de la flora bacteriana vaginal.** (A, B) Imagen típica de una flora vaginal normal, en la cual se aprecia una célula epitelial rodeada de *Lactobacillus*, microorganismos protectores presentes en un tejido sano. (C, D) Imagen típica de una flora vaginal que muestra un inicio de infección bacteriana. Los *Lactobacillus* se van perdiendo, y aparecen

bacterias patógenas con una morfología distinta. (E, F) Imagen típica de un tejido con vaginosis bacteriana. Los *Lactobacillus* están ausentes, y en su lugar una densa colonia de bacterias patógenas rodea a la célula epitelial.

Lactobacillus		Gardnerella / Bacteroides		Mobiluncus	
Cantidad	Puntuación	Cantidad	Puntuación	Cantidad	Puntuación
4+	0	0	0	0	0
3+	1	1+	1	1+ / 2+	1
2+	2	2+	2	3+ / 4+	2
1+	3	3+	3		
0	4	4+	4		
<b>Puntaje Total: <math>\geq 7</math> indica VB, 4-6 Estadio intermedio.</b>					

**Figura 2.** Score de Nugent para la Tinción de Gram en hisopados vaginales

### 3.2. Epidemiología: estadísticas en el mundo y en el Perú.

Un estudio realizado por los Centros para el Control y Prevención de Enfermedades (CDC), con 21 millones de evaluadas, encontró una prevalencia de 29.2%, donde solo el 15.7% de ellas presento síntomas. Esto es de cada 10 mujeres aproximadamente 3 de ellas tienen VB, y de las que poseen VB, solo el 50% son sintomáticas (31). Es importante mencionar que la prevalencia puede variar entre 5 y 51% dependiendo de la cultura o condiciones demográficas (32). En nuestro país se realizó un amplio estudio(15) en algunas ciudades de zonas rurales de Perú en el año 1997 – 1998 bajo una muestra de 752 mujeres evaluadas se obtuvo una prevalencia de VB de 40.8%, en mujeres de edades de 18-67 años. Además, Jones y col. (10) hallaron mediante un estudio reciente una cifra de 27%, en un estudio transversal realizado en varias ciudades de la costa, que mostraron una asociación a la coocurrencia de enfermedades de transmisión sexual (ETS). Entre los factores de riesgos más frecuentes se mencionan la

raza, el tabaquismo, consumo de alcohol, un nivel socioeconómico y/o de instrucción bajos, entre otros (15).

### **3.3. Situación de la VB en el Perú problemática de la VB en la salud pública.**

La VB expone a las mujeres a graves riesgos como son: [1] Mayor susceptibilidad a HIV(7) [2] Mayor posibilidad de transmitir HIV(33) [3] aumento de infecciones post-quirúrgicas(29); [4] aumenta la susceptibilidad a ETS(1), aumentando así la posibilidad de contraer cáncer de cérvix ((34),(14),(35)). [5] Pueden generarse infecciones en el útero y las trompas de Falopio, lo cual provoca infertilidad y embarazo ectópico ((30),(36)). La VB, afecta el bienestar del binomio madre niño generando complicaciones durante la gestación como: [1]Aborto espontáneo((2)) [2]Amenaza de parto pretérmino (18) [3]Parto prematuro(23) [4]Ruptura prematura de membrana (RPM) (4) [5]Infección intra-amniótica(26) [6]Endometritis postparto (26)(29) [7]Infecciones postcesárea (29),[8] Morbilidad ginecológica (37). Por la importancia y la trascendencia de cada una de las posibles complicaciones debidas a una infección por VB se concluye la importancia de contar con un diagnóstico efectivo y al alcance de la mayoría de centros de salud. Debido a que no existen medios suficientes ni médicos patólogos para el diagnóstico en los lugares remotos del país, y a la carencia de algún método cuantitativo y objetivo su aplicación se ve limitada. Esta falta de diagnóstico pone en riesgo la salud del binomio madre niño durante la gestación y provoca problemas relacionados con la mujer en etapa reproductiva en general.

### **3.4. Tratamiento de la VB.**

El tratamiento de VB en una mujer no gestante, beneficia a la mujer aliviando los síntomas y signos de la infección y reduciendo los riesgos de complicaciones debido a una coocurrencia con alguna infección (38).

Además, se disminuye el riesgo de otras ETS.

El tratamiento para pacientes ginecológicas se basa en el metronidazol, y la clindamicina(39).

Las recomendadas dosis son:

- Medicación Vía Oral: Metronidazol, 500 mg, dos veces al día, por el lapso de 7 días, o
- Medicación Intravaginal: Metronidazol gel, 0,75%, un aplicador lleno (5 g), una vez al día, por el lapso de 5 días, o
- Medicación Intravaginal: Clindamicina crema, 2%, un aplicador lleno (5 g), al acostarse, por la cantidad de 7 días.

De preferencia se debe evitar el consumir bebidas alcohólicas durante el periodo de tratamiento y hasta 24 horas después de terminarlo.

Además, existen regímenes alternativos:

- Medicación Vía Oral: Clindamicina, 300 mg, dos veces al día, durante 7 días, o
- Medicación Intravaginal: Clindamicina óvulos, 100 mg, antes de acostarse, por 3 días.

El porcentaje de curación a la semana de tratamiento es de 80% a 90%, aunque existe una tasa de recurrencia a los tres meses que alcanza a 15% o 30%. La recurrencia es por reinfección, principalmente asociado a la actividad sexual.

### **3.5. Reconocimiento de patrones y sus aplicaciones en diagnóstico de enfermedades telemedicina y telediagnóstico.**

El estudio de reconocimiento de patrones aplicado a la biología con fines de diagnosticar algún tipo de enfermedad es relativamente reciente. Comanicui (40) para poder hacer la discriminación y selección de linfomas malignos de linfocitos crónicos de leucemia crearon un sistema de recuperación de imágenes. Con un fundamento de distribución Gaussiana se realiza el análisis multivariado como en los trabajos realizados por Sampat (41) en la clasificación de imágenes de cromosomas; Botigli(42) en el reconocimiento de colonias celulares y por ultimo Trattner (43) que identifico con éxito diversos tipos de bacterias. Ross (44) buscó seleccionar y diagnosticar especies del genero Plasmodium. Finalmente, en el 2007, Castañon (45) Desarrolló un sistema de diagnóstico en línea de Eimeria el cual también se basó en el análisis de imágenes digitales. En la actualidad no existe un modelo cuantitativo y objetivo, para el diagnóstico de VB en base a las fotografías de las muestras e identificación de morfotipos involucrados. Nuestro equipo propone en este proyecto el desarrollo de un software para el diagnóstico automático de VB y poner en línea un sistema de telediagnóstico disponible para su uso por cualquier centro de salud.

En los últimos años, nuestro equipo ha desarrolló un programa computacional en el reconocimiento automático de patrones para el diagnóstico de TBC

en cultivos MODS (microscopic-observationdrug-susceptibility) desde fotografías digitales (46). Este algoritmo de reconocimiento de patrones se basa en la cuantificación de parámetros geométricos posee una sensibilidad de 99.7% y una especificidad de 99.4%. Además, nuestro grupo desarrollo con éxito un software para el reconocimiento automático de parásitos intestinales, mediante el análisis geométrico y de iluminación de fotografías digitales tomadas a muestras de heces. Los resultados han demostrado poder diagnosticar con una sensibilidad promedio de 98% y una especificidad promedio de 98% para cada parasito en el estudio. Para VB nuestro grupo ha explorado el uso de herramientas computacionales para su diagnóstico, el procedimiento que hemos seguido involucra un estudio piloto preliminar de procesamiento de imágenes (Ver Apéndice-VB1), seguido de la exploración de una serie de parámetros geométricos. Las distribuciones estadísticas de estos parámetros geométricos se obtuvieron tanto para las bacterias 'dañinas' como para las bacterias 'beneficiosas' (Ver Apéndice-VB2). Con el estudio preliminar realizado incrementamos las posibilidades de éxito del presente estudio. El siguiente paso que desarrollamos, consiste en evaluar nuestro estudio preliminar encontrando un tratamiento de imágenes ad-hoc e identificando las mejores características geométricas capaces de predecir VB, obteniendo así software automático para el diagnóstico para muestras.

#### **4. HIPÓTESIS INICIAL**

Se puede diagnosticar VB mediante un algoritmo automático basado en la identificación de diversos morfotipos en las muestras de hisopados vaginales mediante el reconocimiento de patrones, logrando alcanzar una sensibilidad de 95% y una especificidad del 90%, evaluada en base a campos de la muestra seleccionados por un técnico de salud.

## **5. METODOLOGÍA.**

### **5.1. Diseño de estudio**

El diseño del estudio es **Evaluación de prueba diagnóstica**, tomando como prueba a evaluar el software automático SoftWarmi para el diagnóstico de VB en láminas de hisopados vaginales.

### **5.2. Población y muestra.**

**Población Objetivo:** 59 láminas de hisopados vaginales de entrenamiento de la Universidad de Washington (UW) en el diagnóstico de VB (Figura 3).

**Datos de la Muestra:** 59 láminas de hisopados vaginales usados en la enseñanza del laboratorio de E.T.S. de la Universidad de Washington provistas por el MD. PhD. King K. Holmes para el estudio, dichas láminas cuentan con un diagnóstico y son anónimas.

#### **Tamaño de Muestra**

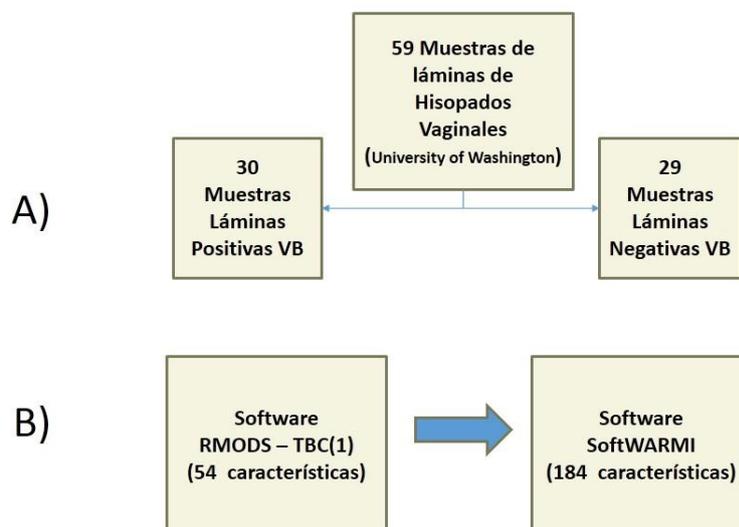
El set de láminas para entrenamiento de la Universidad de Washington tiene 59 muestras. La clasificación de las láminas recibidas es 30 con diagnóstico positivo a VB y 29 negativas para VB.

### **5.3. Procedimientos para desarrollo de SoftWarmi.**

Nuestro laboratorio investigo previamente sobre el diagnóstico automático de TBC en base de fotografías de muestras de cultivos MODS. Este diagnóstico se basó en una versión del algoritmo de extracción de características y que se publicó en el año 2013(1). Esta la versión del software automático se evaluaron y calcularon 54 características. Durante la programación y el proceso de

extracción de características para SoftWarmi, se han extraído 184 características de cada objeto, lo que significa la evaluación y cuantificación de 130 características nuevas. Además se implementó un procesamiento de imagen para la fotografías de muestras de hisopados vaginales. También se han generado modelos estadísticos para la clasificación de objetos, fotos y muestras (Figura 3).

Para cada uno de los objetos segmentados se calcularon 184 características cuantitativas luego se procedió a crear una base de datos y un modelo estadístico de predicción como se explica más adelante, aquí damos una información detallada de los procedimientos realizados, el presente programa automático SoftWarmi estuvo basado en el algoritmo anterior, pero se adaptó optimizo y se crearon características cuantitativas específicas para este fin, además de un procesamiento de imágenes ADOC.

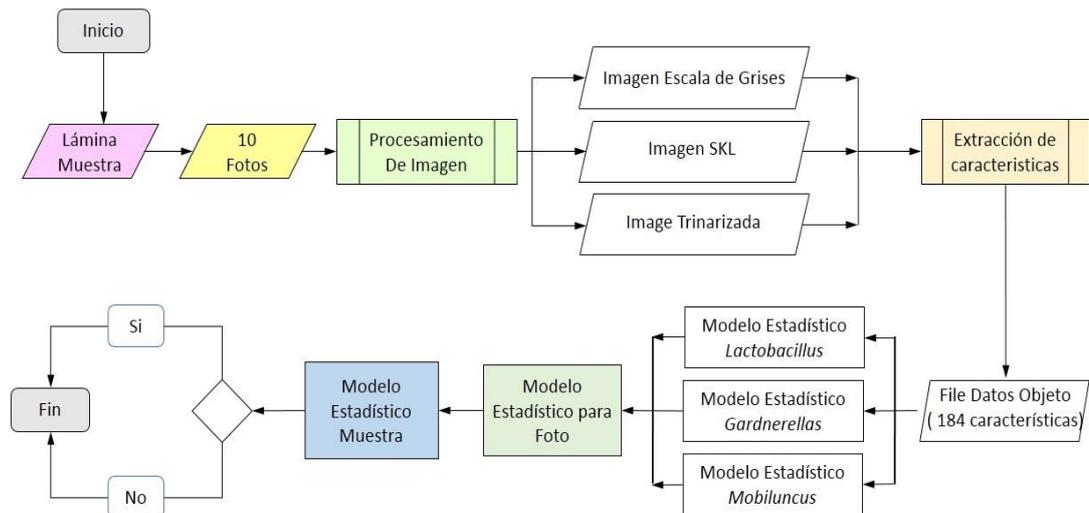


Lenguaje de Programación C.  
 Tiempo Aproximado de procesamiento por foto: 35 segundos.

**Figura 3.** Generalidades. A) Clasificación de las muestras de entrenamiento. B) Mejora de algoritmo de extracción de características cuantitativas para objetos en una fotografía microscópica.

En este proyecto, se propuso el desarrollo y validación de un software para el diagnóstico automático de vaginosis bacteriana, este diagnóstico se realiza en base al conteo de bacterias asociadas a que permitirá determinar cuantitativamente el nivel de infección de vaginosis bacteriana de acuerdo con la score de Nugent, incluyendo específicamente a los morfotipos *Lactobacillus*, *Gardnerella* y *Mobiluncus*.

A continuación detallaremos en general el proceso de desarrollo del software que se explicara en detalle en las siguientes secciones. En la Figura 4 se observa que empezamos con la obtención de las fotografías (10 fotos por lámina) de las muestras de hisopados vaginales, las muestras han sido proporcionadas por la Universidad de Washington. A cada una de las fotografías obtenidas se les ha realizado un procesamiento de imagen que brindo como salida 3 archivos de imagen: La fotografía en escala de grises, la imagen de los esqueletos de los objetos segmentados y por último la fotografía trinarizada. Estas tres imágenes son los archivos de entrada del proceso de extracción de características, que termina con un archivo de datos cuantitativos y categóricos de cada objeto segmentado de la fotografía. Cada objeto es evaluado en base a los tres modelos de clasificación de morfotipos diseñados: el modelo Estadístico para *Lactobacillus*, *Gardnerellas* y *Mobiluncus*. Luego se procede al conteo de los objetos positivos a cada modelo en cada fotografía, se procesa con el modelo Estadístico diseñado para clasificar las fotos con vaginosis bacteriana. Al obtener el resultado de las 10 fotografías de la muestra se calcula el valor del modelo para muestra, que concluye con los resultados del diagnóstico por muestra procesada y se finaliza el proceso completo.



**Figura 4.** Diagrama de Flujo del proceso general del Diagnóstico automático de VB: Inicio. Ingresar la muestra. Digitalizada (10 fotos por lámina). Cada foto ingresa al procesamiento de imagen. Salida: Imagen en escala de Grises, imagen SKL, imagen Trinarizada. Se segmenta cada foto en objetos (1500 por foto aprox.). Cada objeto ingresa al proceso de extracción de características (184 características) e ingresa al modelo de objetos. La foto es evaluada por el modelo de foto. Las 10 fotos son evaluadas según el modelo para muestra. Se emite un diagnóstico de VB por lámina.

### 5.3.1. Obtención de fotografías.

El programa desarrollado buscó diagnosticar VB, mediante el conteo de los diferentes morfotipos presentes en láminas de hisopados vaginales con tinción Gram.

#### Sistema de adquisición de imágenes

Cabe resaltar las condiciones que se estandarizaron para la colección de imágenes de dicha prueba:

- La iluminación (nivel de luz y rejilla).
- El microscopio Directo Marca Zeiss.
- El zoom del lente del microscopio 100x.

- El zoom de la lente de la cámara 10x.
- Cámara mini VID – LWScientific de 3.2 Megapíxeles.
- Laptop Toshiba (Windows 7 – Ram 2 Gb).
- Aceite de inmersión.

A partir de fotos tomadas con estas condiciones de estandarización comenzamos nuestro trabajo.



**Figura 5.** Sistema de digitalización de láminas de hisopados vaginales.

Para evaluar el software de diagnóstico SoftWarmi se procedió con la adquisición de fotografías de las muestras. Con este fin se trabajó con un Ing. Biotecnólogo Ing. Andres Gutierrez, entrenado que realiza la selección de campos válidos para el diagnóstico de VB. Las fotografías se capturarán por el técnico de salud. De cada lámina se obtendrán 10 campos de lectura (10

fotografías). Estos campos y/o fotografías serán comparados y evaluados por diferentes diagnósticos que se numeran a continuación:

- Cada lámina proporcionada por la universidad de Washington posee un score, el cual es nuestro Gold standard. El primer Score considerado en el análisis será el Gold standard.
- El técnico de salud calculará un score propio que dependerá de los 10 campos, como en la lectura que realiza normalmente elegidos según el protocolo del Score de Nugent. A este diagnóstico lo llamaremos segundo Score.
- El técnico de salud, al momento de identificar cada campo para su diagnóstico procederá a fotografiar el campo. Al final de su procedimiento tendrá 10 fotografías de los 10 campos considerados por él. Estas 10 fotografías serán evaluadas por el software el cual procesará su diagnóstico mediante el promedio de los 10 score de cada foto. A este diagnóstico lo consideraremos como el tercer diagnóstico.

En conclusión, el software SoftWarmi evalúa la muestra en base al score basado en las 10 fotografías de los campos del técnico de salud entrenado.

### **5.3.2. Procesamiento de imagen.**

**Objetivo:** Obtener una mejoría en el contraste de la iluminación de la fotografía y así alcanzar una máxima definición en la forma topológica de los objetos digitales llamados *Lactobacillus*, *Gardnerella* y *Mobiluncus* además de eliminar el ruido (manchas, sombras, objetos microscópicos).

**Proceso:** En esta sección intervienen elementos de tratamiento de imágenes propiamente y se realiza la acción en toda la fotografía píxel a píxel. Para cada

paso hemos usado el programa en Lenguaje C que es de licencia gratuita bajo un entorno de Ubuntu-Linux. Los pasos realizados son:

### **Mejora Contraste**

Para resaltar en la imagen los objetos de análisis resaltando las variaciones de luminosidad para una mejor interpretación y segmentación de los objetos aplicamos un contraste dependiente del nivel de iluminación de la fotografía [m17]. Utilizamos el método de expansión del contraste que estira el rango de valores del histograma [m16-m17]. Este procedimiento depende del histograma de la imagen [m18].

### **Conversión Escala de Grises**

Una fuente de luz provoca diferentes efectos en diversos ángulos que se deben eliminar de la imagen, así como también algunos factores externos que afectan la imagen en diferentes rangos de colores. El proceso de eliminación se realiza mediante la transformación de la imagen de colores e intensidades con el cambio de RGB a escala de grises. El proceso toma la iluminación monocromática parametrizándolas con los valores estándar NTSC, para esto se aplican los coeficientes con valores que tienen relación con la sensibilidad del ojo humano a los colores RGB.

$$I = 0.2989*RGB\_img(:, :, 1) + 0.5870*RGB\_img(:, :, 2) + 0.1140*RGB\_img(:, :, 3)$$

La escala de grises varia en una escala de 256 valores, desde 0 al 255. Se ha normalizado esta escala con lo que varían entre [0,1].

### **Filtro Gamma**

Para ajustar los valores de los pixeles de la imagen respecto una mínima diferencia de color y aclarar la imagen de acuerdo al valor de cada pixel y al

ancho de cada bit utilizamos este filtro. Este filtro permite dedicar la cantidad de pixeles justos a los valores con mayor o menor luminosidad dependiendo de los valores que tome. Para el algoritmo se eligió el valor de 1.1, valores mayores de 1 permiten aligerar la imagen, iluminarla levemente para evitar contactos inexistentes dependiendo de la densidad de la muestra (m30, m31).

### **Binarización Adaptativa**

Con la finalidad de segmentar adecuadamente los objetos del fondo de la muestra uno de los procesos que se buscan es la binarización local, con lo que se adapta a los diferentes niveles de iluminación en sección de las fotos de la imagen. De entre los métodos de binarización local existentes se eligió el algoritmo de Niblack que adapta el índice del umbral para la binarización dependiendo de las características de la vecindad local de cada pixel. Inicialmente se pensó en un umbral binarización global, pero al analizar los diferentes tipos de iluminación dentro de cada fotografía de la muestra se procedió al análisis de métodos de binarización local de ellos se optó por el método de Niblack dado que nuestro microscopio posee un sistema de iluminación no uniforme que depende de la luz y de la muestra fotografiada.

### **Suavizar siluetas de los objetos**

Con el objetivo de suavizar los contornos de los objetos dentro de las fotografías de las muestras, aplicamos este filtro. Entre los procesos del filtro está el eliminar los pixeles con intensidad diferente a la de los pixeles en la vecindad del punto específico. El objetivo es eliminar los pixeles cuya intensidad influyen en la resolución de la imagen, esto se consigue aplicando un filtro llamado Mediana. Así podemos atenuar ruidos llamados sal y pimienta, que son puntuales o

granulares en la fotografías , y esto se realiza cuando la fracción de píxeles con ruido en su vecindad es menor que la mitad de los valores de la intensidad de los vecinos del pixel analizado.

### **Etiquetado**

Esta función permite el etiquetado de cada objeto segmentado en la imagen después del procesamiento de la misma, las etiquetas se crean basado en los píxeles conectados entre sí en la imagen binaria y en la identificación de los píxeles no-nulos. El etiquetado es realizado a la imagen binarizada con ella se reconoce cada objeto segmentado, identificándolos mediante su posición para posteriores funciones. (m32)

### **Exclusión de elementos del borde exterior de la imagen**

Para eliminar los objetos que no están incluidos en la imagen cuya intersección con el borde ( $i = 1, i = \text{ancho}, j = 1, j = n$ ) es no vacía. Primeramente se etiquetan las regiones en la imagen luego se identifican aquellos que están en contacto con los bordes y ellos son excluidos del análisis. Este proceso nos ayuda a eliminar a la vez objetos producidos por la iluminación como sombras de mayor intensidad.

### **Abertura de la imagen (Erosión - Dilatación)**

Con el fin de eliminar ruidos en la imagen, mejorar el contorno de los objetos modificándolos en función de su topología se aplicaron dos operadores morfológicos en conjunto, la dilatación y la erosión en el orden respectivo; es decir el cierre de la imagen. La dilatación estrecha las regiones y hace que los picos sean mayores y más suaves, en cambio la erosión estrecha los picos y

agranda las regiones. Para ambos procesos utilizamos el elemento estructural de orden 3.

### **Filtro de llenado de Huecos**

Para eliminar los huecos en la imagen binaria es decir regiones blancas incluidas en los objetos negros se usó un proceso de llenado de agujeros.

### **Esquelenotización**

Con la finalidad de obtener una representación de la forma que conserve características geométricas de cada objeto en la fotografía de la muestra se procederá al proceso de esquelenotización, (i.e. los esqueletos de cada uno de los objetos de la imagen), este proceso se basa en el adelgazamiento sucesivo de cada objeto en la imagen conservando los principales características topológicas, con la finalidad alcanzar la mejor aproximación a su silueta dada por una curva digital continua de ancho mínimo, esto nos permitirá describir la estructura de cada objeto. Se define el esqueleto de un objeto digital de la imagen binaria como aquel lugar geométrico formada por los centros de las circunferencias maximales inscritas en el borde inicial de dicho objeto, los esqueletos son calculados usando usando la métrica euclidiana.

### **Bordes en la imagen**

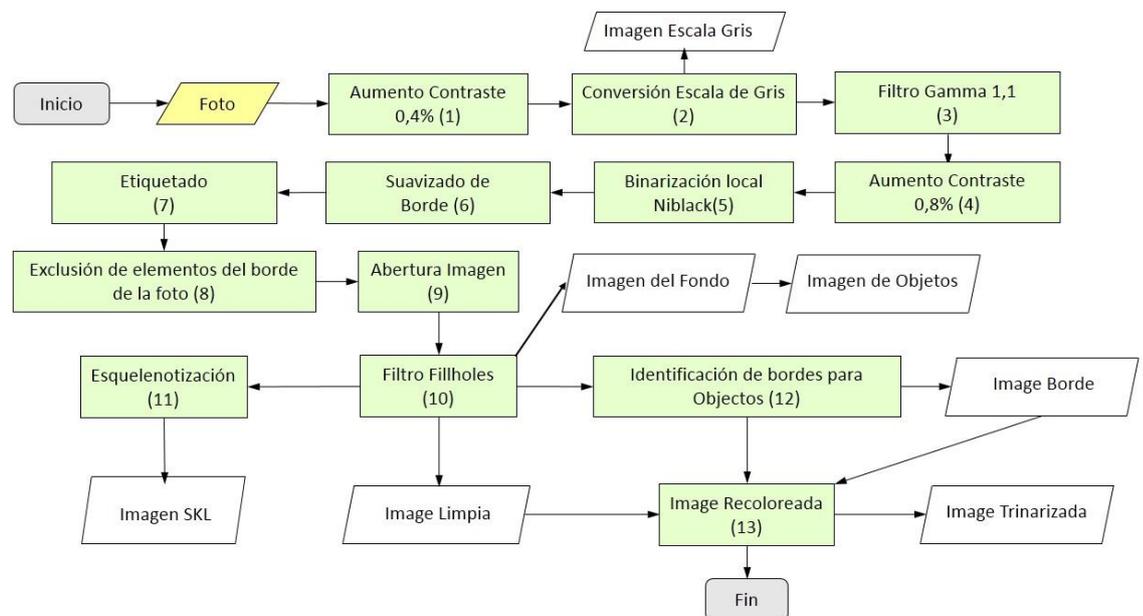
Cuando se analizan los objetos, los bordes de los mismos tienen mucha información de la forma del objeto. Es importante encontrar el algoritmo adecuado que permita mantener el borde del objeto lo más fiel al real con la mínima pérdida de información. El proceso que se aplicará será mediante el estimador gradiente de Sobel. Los algoritmos para encontrar los bordes utilizan

maskaras horizontales y verticales, esto permite extraer la información independiente de la orientación del objeto. La imagen que se obtiene del proceso con el gradiente de Sobel crea un contorno omnidireccional de los objetos en la imagen original.

**Trinarización** Para poder trazar hacer el análisis del espesor y el nivel de luminosidad requerimos una imagen trinarizada. Esto es diferenciar con color el fondo (gris), el borde (negro) y el interior (blanco) de cada objeto en la imagen.

### Modelo del Procesamiento de Imagen (en Fotos)

El tratamiento a la imagen ha ido modificándose con el fin de resaltar y definir los objetos de nuestro interés, esto es, poder obtener un buen contraste y una silueta “suave” de los objetos llamados *Lactobacillus*, *Gardnerella* y *Mobiluncus*. El resultado de este proceso se puede visualizar mejor en el siguiente cuadro



**Figura 6.** Diagrama de Flujo del procesamiento de imagen.

### **5.3.3. Extracción de características.**

**Objetivo:** Extraer cada característica geométrica, cuantificarla y analizarla desde el punto de vista de coherencia biológica y estadística para en vista de su significancia y correlación elegir las mejores para generar un modelo estadístico apropiado para el diagnóstico de cada morfotipo relacionado.

**Proceso:** En esta sección intervienen elementos de geometría digital y euclidiana, topología digital, medidas de textura, modelación gaussiana, transformada de Fourier y calculo diferencial, entre otros.

Y se han creado algunas medidas en vista de la necesidad de extraer características que permitan simular el reconocimiento realizado por los expertos. Para cada paso hemos usado lenguaje C, que es de licencia gratuita y trabaja bajo Linux, con algunas librerías especializadas en el procesamiento de imágenes y extracción de características numéricas. Los pasos realizados son:

#### **Identificación de Puntos Críticos**

Para comenzar a analizar el esqueleto de un objeto necesitamos un punto en el, encontrando el primer píxel negro en la imagen dado por un orden lexicográfico, a partir del recorreremos el esqueleto en sentido anti horario hasta llegar al primer píxel que posea solo 1-vecindad y lo llamamos primera punta.

Es importante encontrar la primera punta para asegurarnos llegar a ella solo una vez más, así el recorrido del esqueleto termina al encontrarla llegar nuevamente.

### **Baricentros, puntas, brazos y tronco**

Para transformar el esqueleto en un grafo necesitaremos clasificar los puntos del mismo. Por esto desde la primera punta recorreremos la estructura en sentido anti horario, representando el esqueleto por una matriz de  $2 \times n$ , donde  $n$  es el número de píxeles del mismo.

Al ingresar cada píxel al esqueleto lo analizamos, clasificándolo en tres grupos: puntas, aquellos que solo tienen 1-vecindad; baricentros aquellos que poseen a partir de 3-vecindad. Medios, los que no son ni puntas ni baricentros.

Los medios serán parte del camino digital que conecta baricentros y/o puntas.

Los baricentros son agrupados en clases dependiendo de su proximidad.

- Al camino digital de cada punta a un baricentro se le llama brazo.
- Al camino digital que une dos baricentros se le llama tronco.

### **Transformación de esqueleto a grafo no dirigido**

Para transformar el esqueleto a un grafo necesitamos identificar los vértices y los arcos. Cada punta y baricentro se define como un vértice, los caminos digitales de brazos y tronco son llamados arcos cuyos pesos están definidos como el número de píxeles que los conforman.

### **Extracción de ramas y esqueleto principal primario**

Para identificar el esqueleto principal primario debemos extraer las ramas del esqueleto. Dado el grafo definido anteriormente calculamos su diámetro, obteniendo una sucesión de vértices que conforman el esqueleto principal

primario, transformándolo luego en un camino digital lo obtenemos como una sucesión de vértices continua. Para encontrar las ramas, desde cada punta que no pertenecen al esqueleto principal se encuentra la sucesión de vértices una dicha punta al esqueleto principal, al transformarlo nuevamente a un camino digital obtenemos lo que denominamos una rama.

### **Ajuste de elipse**

Para comparar los objetos digitales a una elipse ajustamos el borde (conjunto de puntos 2D) del mismo mediante una función basada en mínimos cuadrados (47) que encuentra el rectángulo de mínima área sobre la cual se inscribe la elipse de mejor ajuste.

### **Ajuste Rectángulo**

Para encontrar el rectángulo de mínima área que se ajuste al objeto se utilizó la técnica de mínimos cuadrados (47), mediante la librería libre Opencv.

### **Momentos Geométricos**

Cada imagen es transformada de sus dimensiones originales del formato RGB, a solo dos dimensiones, es decir una región plana de 2 dimensiones. La manera de relacionar las medidas de orientación, forma, posición y tamaño de cada objeto se da con los momentos, medidas invariantes que se basan en los pixeles de cada imagen(48).

Los momentos de orden  $p+q$  para una imagen  $I(x,y)$  son :

$$M_{p,q} = \sum_{x=1}^N \sum_{y=1}^M x^p y^q I(x, y).$$

Por ejemplo entre las medidas que se obtienen tenemos, que el momento de orden cero es el área del objeto, para una imagen binaria.

### **Ajuste Circular**

Para realizar el ajuste a círculo de cálculo el centroide del objeto, y el promedio de las distancia del centroide, al borde del objeto.

### **Extracción de borde**

Para extraer el borde del objeto en la imagen, desde una punta del esqueleto primario trazamos una recta vertical para encontrar el primer píxel que tenga un vecino del color del fondo, que es considerado el primero del borde, siguiendo un proceso de iterativo en la dirección de la mano derecha y con una topología digital 8-vecindad obtenemos al final la matriz ordenada del borde de orden  $2 \times n(B)$ , donde  $n(B)$  es el número de píxeles del borde.

### **Extremos del esqueleto**

Para poder encontrar los extremos del esqueleto primario principal solo elegimos el primer y el último píxel del ordenamiento.

### **Procedimiento de convergencia – nuevos vértices**

Para completar el esqueleto principal primario (i.e. donde su píxel inicial y final pertenezcan al borde de la imagen) creamos y aplicamos el procedimiento convergencia. Dicho procedimiento ubica sobre el borde del objeto digital un punto que será el final del esqueleto primario, es decir lo completa para poder tener un esqueleto final que intercepte al borde. Se basa en el algoritmo de puntos interiores de optimización. Para esto definimos la proyección de un punto, sobre el lado opuesto del esqueleto, para dicha sección.

#### ***Algoritmo convergencia***

*x*: píxel extremo del esqueleto principal primario de la imagen

*M\_m*: conjunto de valores de diferentes pendientes

*Inicio1* (input *x*)

*Inicio 2* → para cada pendiente *m\_i* en *M\_m*

*Trazo de recta* (*m\_i*, *x*)

*Intersecciones de recta con el borde* (*x1\_mi*, *x2\_mi*)

*Distancia* (*x1\_mi*, *x2\_mi*) → *Fin 2*

*Mínimo de todas las distancias*

*Ubicación* = *media*(*x1\_mi*, *x2\_mi*),

*Proy* = *proyección de ubicación sobre el borde*,

*Empiezo nuevamente con inicio1*

*Hasta que mínimo de distancias sea mayor a la anterior*

*Fin1*

Desde el punto extremo del esqueleto primario trazamos rectas con diferentes pendientes de las cuales seleccionamos aquella que posea la menor distancia, es decir obtenemos un output de dos ubicaciones de píxeles en el borde, con ellas calculamos la media aritmética teniendo como resultado una nueva ubicación a este píxel procedemos a calcular su proyección sobre la porción de borde opuesto a él. Repitiendo el procedimiento anterior hasta que la nueva distancia obtenida entre estos dos puntos sea mayor a la última considerada.

### **Esqueleto final**

Para terminar de completar el esqueleto primario desde cada extremo del esqueleto inicial aplicamos el procedimiento convergencia a partir del cual obtenemos un punto que colocamos como el extremo de un nuevo esqueleto, desde el trazamos una camino digital recto al último punto del esqueleto primario convirtiendo el nuevo camino en el esqueleto principal y final.

### **Espesor y nivel de iluminación**

Los siguientes procedimientos son usados para dividir el interior del objeto seleccionado en porciones representadas por rectas que usaremos como la data de prueba para establecer el espesor medio del objeto y su nivel de iluminación que nos permita medir la birrefringencia.

### **División del esqueleto principal**

Para dividir el esqueleto principal tomamos puntos en él. Cada uno de ellos igualmente espaciados en el esqueleto principal (cada 5 ubicaciones).

### **Extremos de cada división**

Para encontrar las rectas incluidas en el interior sobre las cuales se realizará el análisis del espesor y de la birrefringencia, desde cada punto que pertenece a la división dada en la sección anterior, trazamos rectas perpendiculares al esqueleto final y hallamos los dos puntos de intersección de ella con el borde del objeto, obteniendo de ellos el segmento incluido en el borde como camino digital.

### **Análisis del espesor**

Para analizar el espesor, tomamos las distancias del paso anterior formando un vector que tenga correspondencia con cada uno de los puntos de la división del esqueleto principal y procedemos a encontrar el máximo espesor, el promedio y la desviación estándar.

### **Análisis de la Refracción - base media foto**

Para analizar la iluminación en cada segmento de recta trazada desde cada división del esqueleto principal incluido en el objeto eliminamos cuatro píxeles de los extremos, porque contienen información de la iluminación del exterior del objeto y no de la refracción que está ubicada en el interior del segmento, después de esta eliminación encontramos el máximo y mínimo valor en la imagen de escala de gris existente y

Verificamos si la diferencia entre ellas es mayor a un nivel de iluminación establecida en diferentes niveles (8, 10, 12, 15, 18, 20 y 22) que dependen exclusivamente de la media de los valores de luminosidad de todos los píxeles de la foto. En caso se cumpla del máximo valor obtenido verificamos cuantos de los píxeles cumplen con una iluminación significativa (según la escala) y lo expresamos en función del porcentaje del segmento, teniendo así una medida de la iluminación y con ella la refracción que depende del nivel de iluminación por camino estudiado y de la longitud del segmento analizado en ese procedimiento.

### **Análisis de la Refracción - base mascaré de color R, G, B**

Para analizar la iluminación en cada segmento de recta trazada desde cada división del esqueleto principal incluido en el objeto eliminamos cuatro píxeles

de los extremos, porque contienen información de la iluminación del exterior del objeto y no de la refracción que está ubicada en el interior del segmento, después de esta eliminación encontramos el máximo y mínimo valor en la imagen de escala de mascarará de color existente y

Verificamos si la diferencia entre ellas es mayor a un nivel de iluminación establecida en diferentes niveles (8, 10, 12, 15, 18, 20 y 22) que dependen exclusivamente de la media de los valores de luminosidad de la foto en la respectiva mascarará de color de todos los píxeles de la foto. En caso se cumpla del máximo valor obtenido verificamos cuantos de los píxeles cumplen con una iluminación significativa (según la escala) y lo expresamos en función del porcentaje del segmento, teniendo así una medida de la iluminación y con ella la refracción que depende del nivel de iluminación por camino estudiado y de la longitud del segmento analizado en ese procedimiento.

### **Análisis de la forma**

En esta sección analizamos la forma estudiando la variación que experimenta el esqueleto principal, en la curvatura y respecto a la línea de regresión de los puntos que lo conforman.

### **Línea de Regresión**

Para evaluar la proximidad a una recta del esqueleto principal final calculamos estadísticamente la línea de regresión con los puntos que lo conforman, luego encontramos el cuadrado de la distancia y la desviación estándar desde cada punto del esqueleto principal, para poder medir de esta forma tanto se desvía de esta línea de regresión el esqueleto.

## **Curvatura**

Para medir un grado de sinusidad es necesario el valor de la curvatura para ello el esqueleto principal se modela mediante la función gaussiana y aplicando la transformada rápida de Fourier(49), hallamos la curvatura de cada punto incluido en el.

## **Análisis de Onda (cantidad-curvatura – tamaño de onda)**

Para cuantificar el grado de sinusidad se ha analizado el número de ondas y el tamaño de cada una de forma individual. Estadísticamente se encontró un punto de curvatura mínima a partir de lo cual consideramos que se comienza a formar una onda y se mide el número de píxeles en ella finalizando cuando se alcanza el nivel inferior de la curvatura. Como salida de este procedimiento obtenemos: el tamaño de la onda, el número de ondas, la curvatura y la proporcionalidad directa e inversa entre ellas.

## ***Parámetros Secundarios***

### **Extracción de Objetos**

Para hacer el estudio individual de cada objeto es separado de la imagen original.

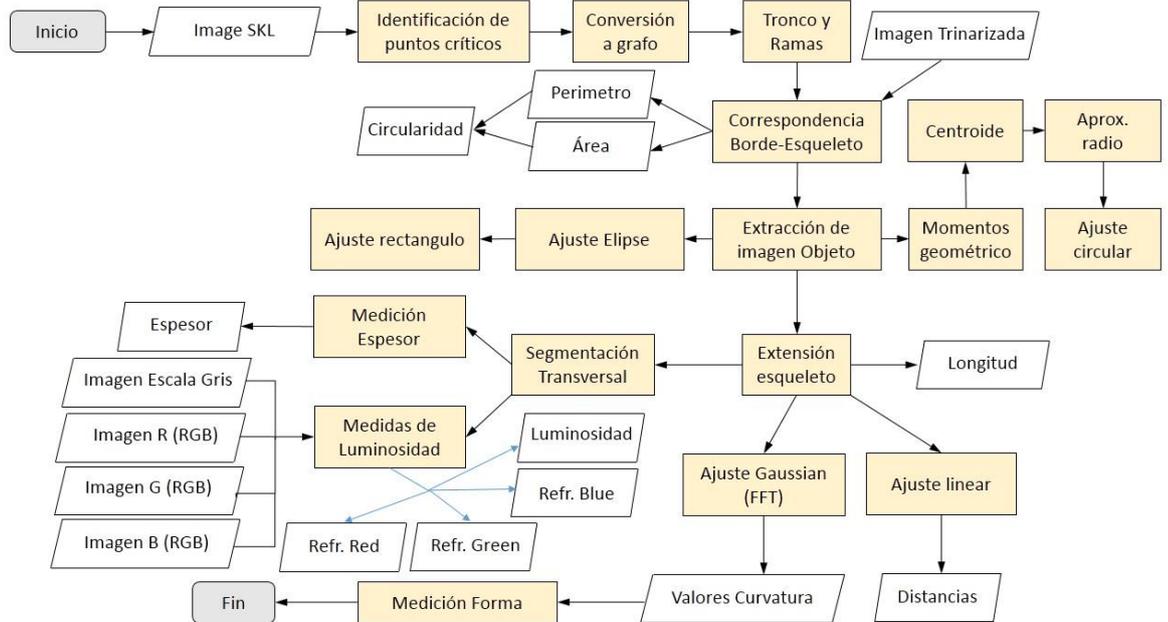
### **Circularidad – Área**

Para medir el grado de semejanza del objeto a un círculo se calcula el parámetro de circularidad como la relación entre el cuadrado del perímetro y el área(50).

Mientras más cercano a uno más parecido a un círculo.

$$***Circularidad = 4 \pi \text{Área} / \text{perímetro}^2***$$

El gráfico mostramos el diagrama general del algoritmo en la extracción de características de los objetos, podemos analizar el orden de los procesos y la dependencia de algunos con relación a los anteriores.



**Figura 7.** Diagrama de Flujo del proceso de extracción de características.

### 5.3.4. Análisis de clasificación.

La experta Dr. Karol Spiegel realizó una clasificación y diagnóstico de la muestra en tres niveles: Para objetos - Nivel 1, para fotos – Nivel 2, y para muestras – Nivel 3.

La forma natural en la que un técnico entrenado brinda el diagnóstico de VB a una lámina, sigue la misma lógica es por este motivo que a continuación presentamos cada modelo generado en cada uno de los niveles que tendrá como consecuencia el diagnóstico por lámina

En esta sección mencionaremos las herramientas que utilizamos para evaluar los diagnósticos del software en cada uno de los niveles y su relación con el diagnóstico del técnico, así como las medidas utilizadas en evaluación del mismo. .

Detallaremos el plan de análisis en cada caso a continuación:

#### **5.3.4.1. Modelo de objetos – Nivel 1.**

Se elegirá un nuevo modelo estadístico de predicción de objetos para el diagnóstico de cada morfotipo.

Los elementos de análisis serán los objetos obtenidos en cada fotografía. La base de datos completa presenta un número de objetos de cada morfotipo: 289 *Lactobacillus*, 1016 *Gardnerella*, 110 *Mobiluncus* y 215 *Objetos negativos*. De estos objetos han sido seleccionados como positivos aquellos que presentarán características morfológicas determinadas, y hayan sido clasificados por la experta en el diagnóstico. Cada objeto digital constará con un conjunto de 184 variables numéricas, variables del proceso de extracción de características realizada.

Los objetos seleccionados para el análisis se determinaron en base a la selección realizada por la Dra. Spiegel. En la base de datos existían 1630 objetos aprox., con la siguiente clasificación: 289 *Lactobacillus*, 1016 *Gardnerellas*, 110 *Mobiluncus* y 215 restos celulares.

Para clasificar las variables de cada modelo, para cada morfotipo, se realizó un análisis univariado considerando el Pseudo R<sup>2</sup>, el odds ratio y el p-value para discriminar a las variables analizables a incluir en el modelo de diagnóstico, conservando sólo las que presentan un p-valor menor a 0,05.

Se creó un modelo de diagnóstico utilizando una regresión logística múltiple. La regresión logística es el método de análisis adecuado para modelar una variable

respuesta binaria, del tipo presencia o ausencia de morfotipo, que permite el uso conjunto de covariables de tipo categórico y continuo, proporcionando interpretación biológica a sus parámetros.

En el proceso de selección del modelo se tomará cuenta los siguientes aspectos:

Análisis univariado de las variables, selección de escala apropiada para las variables continuas, y consideración de términos de interacción, basada tanto en criterios estadísticos como teóricos.

La correlación entre las variables a incluir no deberá poseer un alto grado de correlación (mayor a 0,75). Con este objetivo la elección de las variables a usar en el modelo se llevará a cabo bajo la consideración de dos criterios: (a) Analizar la matriz de correlación de todas las variables y (b) Clasificar por grupos la correlación y la significancia de cada variable desde el punto de vista estadístico, geométrico, iluminación y color.

Se seleccionará el modelo logístico más representativo, se procederá a analizar la coherencia biológica y estadística de cada variable. Se estimará la eficiencia y la exactitud de la prueba diagnóstica a partir del área bajo la curva ROC. Otro indicador que se usará es el índice J de Youden, que nos brindará una probabilidad de corte para la clasificación de nuestro modelo.

Del modelo logístico se calculará la sensibilidad y la especificidad en el grupo de cada morfotipo utilizados para la creación del modelo.

### **Evaluación del modelo de Objeto - Nivel 1**

Los Diagnósticos (Dx.) que consideraremos para cada lámina serán 2: 1. El Dx. de Dr. Karol Spiegel – Gold standard 2. El Dx. de SoftWarmi-Objeto.

De cada diagnóstico mencionado para cada uno de los morfotipos involucrados, se ha estimado la sensibilidad y especificidad comparadas Gold estándar brindado por la experta Dra. Karol Spiegel.

#### **5.3.4.2. Modelo de fotos – Nivel 2.**

Se crearon la base de fotos en relación a la clasificación realizada por la Dr. Karol Spiegel. Para este modelo se generaron dos bases una para el entrenamiento del modelo (50 fotos) y otra base para la prueba del modelo seleccionado (66 fotos).

Los datos generados para cada foto corresponden al conjunto de datos globales, así como la evaluación de cada modelo de objeto en todos los elementos de las fotografías, cada foto contaba con un conjunto de 189 características.

Entre las características consideradas están de iluminación de la fotografía, la desviación estándar de los niveles de iluminación, el promedio y la desviación estándar del fondo de la fotografía y de los objetos. También se consideraron medidas de similitud para el fondo y los objetos segmentados en cada foto, en base a contraste, luminancia y estructura. Además que se han considerado las mejores probabilidades de los 10 modelos de objetos diseñados. Se diseñaron en total: 4 modelos de objetos para *Lactobacillus*, 3 modelos para *Gardnerella* y 3 modelos de *Mobiluncus*. De cada uno de los modelos mencionados fueron ordenados en base a las 8 mayores probabilidades de cada objeto en toda la foto. También se consideró el promedio de estas mejores probabilidades para los objetos en la fotografía.

#### **Evaluación del modelo de Foto - Nivel 2**

Los Diagnósticos (Dx.) que consideraremos para foto serán 3: 1. El Dx. de Dr. Karol Spiegel – gold standard 2. El Dx. del técnico de salud 3. El Dx. de SoftWarmi-Foto.

El Modelo de fotografía se evaluó con la sensibilidad y especificidad de la prueba sobre dos bases: la base de entrenamiento y la base de prueba.

Se midió el porcentaje de agreement entre el Dx. por foto de SoftWarmi y el Dx. De la experta. Así como también se evaluó el porcentaje de agreement de la experta respecto al Técnico entrenado.

En ambas comparaciones se procedió también a calcular el índice Kappa como medida de comparación.

### **5.3.4.3. Modelo de muestras – Nivel 3.**

Para generar el modelo de muestra se trabajó con una base de 60 láminas, los datos que se asignaron a cada lámina fueron el número de campos sospechosos por muestra, teniendo de cada lámina la evaluación del software de 10 campos. Cada campo fotografiado para cada lámina fue seleccionada por el técnico de salud entrenado para tal fin.

#### **Evaluación del modelo de Muestra - Nivel 3**

Los Diagnósticos (Dx.) que consideraremos para cada lámina serán 2: 1. El Dx. de la lámina – Gold standard 2.El Dx. de SoftWarmi-Muestra, calculado a partir de 10 campos seleccionados por el técnico de salud entrenado.

Se calculó la sensibilidad y especificidad de la prueba teniendo como Gold estándar el Dx. brindado por la University of Washington.

Para medir el grado de concordancia entre el Dx. de la lámina para la University of Washington y SoftWarmi se calcularon también el porcentaje de agreement y el índice Kappa correspondiente.

### **5.3.5. Consideraciones éticas.**

El estudio fue remitido al Comité Institucional de Ética de la Universidad Peruana Cayetano Heredia para su aprobación. Se solicitó la exoneración de la revisión del comité de ética por que los láminas no proviene de pacientes si no de muestras de entrenamiento anónimas de la Universidad de Washington.

El autor declarará no tener conflictos de interés en el desarrollo de esta investigación.

### **5.3.6. Recursos Requeridos**

El proyecto será realizado en los Laboratorios de Investigación y Desarrollo (LID), de la Facultad de Ciencias de la Universidad Peruana Cayetano Heredia. El desarrollo de este proyecto comprende la colaboración de dos participantes colaboradoras: (1) El laboratorio de bioinformática, en donde se desarrollará el software y se desarrollará la plataforma WEB de diagnóstico de vaginosis bacteriana en línea; (2) La Unidad de Epidemiología de La facultad de salud pública obtuvo las láminas de frotis vaginales anónimas necesarios para la optimización del software.

El LID constituye uno de los centros de investigación más importante del país con una alta productividad científica reconocida internacionalmente. El LID está ubicado en un edificio de 3 pisos el cual comprende 41 laboratorios de investigación en diferentes áreas, entre ellos un laboratorio especializado en tuberculosis y un laboratorio de bioinformática. La UPCH cuenta con un moderno ambiente para albergar equipos de cómputo de alto rendimiento, tal como un clúster computacional utilizado para las pruebas. Este ambiente cuenta con sistemas de acondicionamiento de temperatura, humedad y tensión eléctrica, al igual que un

sistema de UPS y un grupo electrógeno propio para garantizar que los sistemas de cómputo operen de manera ininterrumpida. Adicionalmente este ambiente cuenta con el soporte técnico de personal altamente calificado para velar por el buen uso y mantenimiento de los equipos.

## 6. RESULTADOS.

### 6.1. Procesamiento de imagen.

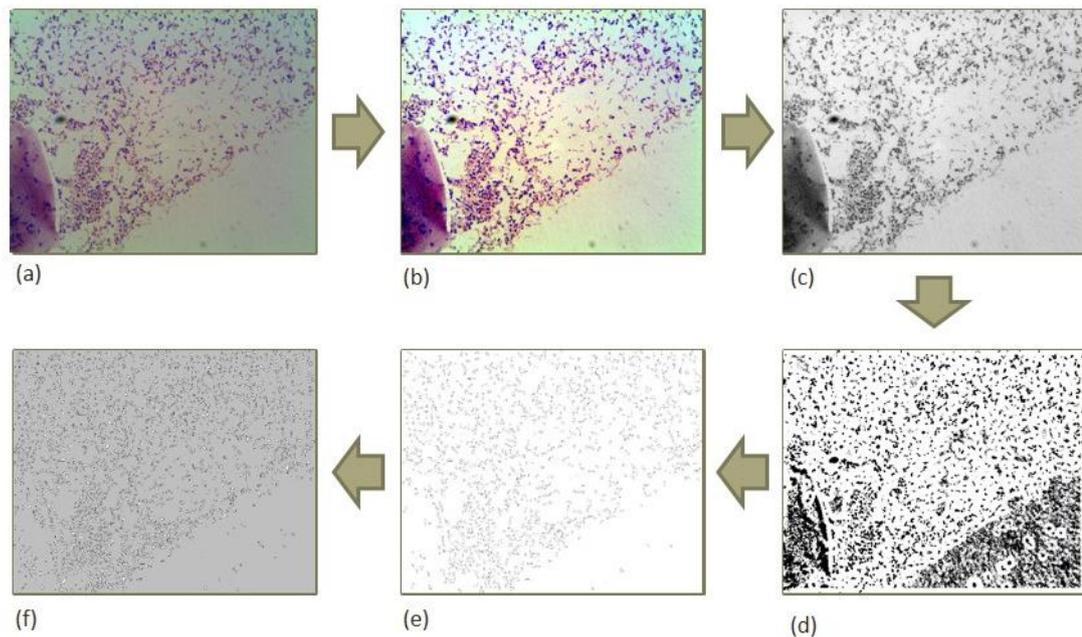
El procesamiento de imagen que se realizó para la identificación de morfotipos que caracterizan la presencia de VB en las láminas de hisopado vaginal. El diagnóstico por el score de Nugent se hace mediante la identificación y el conteo de morfotipos *Lactobacillus*, *Gardnerella* y *Mobiluncus*. Por este motivo el procesamiento de imagen se enfatizó en la segmentación de los morfotipos mencionados, pero sobre todo en un proceso que permita disminuir las uniones entre objetos próximos sin pérdida significativa de su forma. Se han analizado 100 fotografías elegidas al azar al detalle para realizar la identificación de los morfotipos seleccionados por la expertas y han podido ser seleccionados de 1000 a 1500 objetos por fotos aproximadamente, de lo mismo podemos concluir que el procesamiento de imágenes elegido es el adecuado.

Cabe mencionar que se han generado 5 procesamientos de imagen diferentes y se ha elegido el que segmente de mejor manera los morfotipos buscados disminuyendo las conexiones entre objetos próximos.

El procesamiento de imagen es crucial en cualquier proyecto de reconocimiento de patrones basado en imágenes, en este caso se ha dado énfasis a los objetos en una clasificación por área ya que los morfotipos *Gardnerellas* o *Mobiluncus* son de áreas mínimas en relación a otros objetos presentes en la imagen de un campo del hisopado.

A continuación, mostramos un esquema del procedimiento para observar los cambios respecto a los filtros tal como podemos observar en la siguiente figura:

# Procesamiento de Imagen



**Figura 8.** Procesamiento de Imagen: (a) Imagen Original. (b) Imagen con un aumento del contraste de 0.4%. (c) Imagen convertida a escala de grises. (d) Imagen binarizada localmente. (e) Imagen con bordes de objetos segmentados. (f) Imagen Trinarizada (fondo->gris; borde de objetos->negro; interior de objetos->blanco).

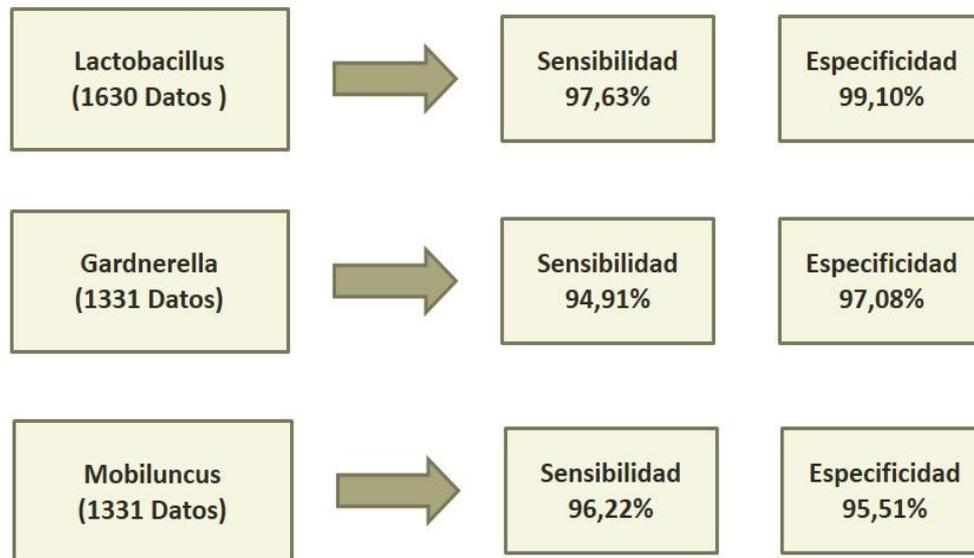
## 6.2. Reconocimiento de patrones.

### 6.2.1. Modelo objetos – Nivel 1.

Al procesar una imagen la división mínima de análisis es un pixel, se utilizan los filtros de imagen para agrupar dichos pixeles por las condiciones de forma, contraste color o iluminación y realizar una segmentación de objetos. Esta es la división sobre la cual se exploran las características identificando los morfotipos involucrados en el diagnóstico de VB (*Lactobacillus*, *Gardnerellas*, *Mobiluncus*).

Cada modelo estadístico para cada morfotipo se calculó en base a las características biológicas y cualitativas del morfotipo seleccionado, agruparemos los resultados por morfotipo analizado.

## Resultados – Modelo Objeto



Lenguaje de Programación C.

Tiempo Aproximado de procesamiento por foto: 35 segundos.

**Figura 9.** Cuadro Resumen de modelos de objetos para morfotipos: *Lactobacillus*, *Gardnerellas* y *Mobiluncus*.

### 6.2.1.1. Modelo *Lactobacillus*.

*Base de Datos:*

La base de datos para *Lactobacillus* se conformó de 1630 objetos: 289 *Lactobacillus* y 1341 objetos negativos conformados por *Gardnerellas*, *Mobiluncus* y restos celulares.

*Variables Significativas*

Se calcularon 184 características para cada objeto *Lactobacillus* seleccionado. En la siguiente tabla se presenta la descripción de las variables significativas involucradas en el modelo estadístico. Entre las características resaltantes se encuentran 4 de forma, de ellas dos representan el comportamiento de los extremos

del objeto (prompuntas, prompuntasredesst), una representa la uniformidad del espesor del objeto (espesordesvstan) y por último una el comportamiento lineal de la forma del objeto (promfor). La iluminación es representada por una variable (bimedia); y el color de los objetos productos de la tinción Gram por tres variables (lpsval, lpvval, bidesv\_r). Ver Tabla.

Característica	Cuantificación	Representatividad
prompuntas	Promedio de las distancias entre los pixeles correspondientes opuestos a cada pixel extremo del esqueleto principal.	Forma.
prompuntasredesst	Promedio de las Desviación estándar de las distancias entre los pixeles correspondientes opuestos a cada pixel extremo del esqueleto principal, relativo a la longitud del objeto.	Forma.
espesordesvstan	Promedio de la desviación estándar de las longitudes de las líneas transversales al esqueleto principal separadas cada tres pixeles a lo largo de la longitud.	Forma.
promfor	Promedio de la distancia del esqueleto principal a la línea de regresión lineal formada los puntos del esqueleto.	Forma.
bimedia	Promedio de los valores de iluminación en escala de grises del objeto.	Iluminación.
lpsval	Valor de saturación(S) , de la escala de color HSV, transformado desde los valores de RGB, de los pixeles internos de las líneas interiores transversales al esqueleto principal. Rango de valores [0,100].	Color.
lpvval	Valor de value(V) , de la escala de color HSV, transformado desde los valores de RGB, de los pixeles internos de las líneas interiores transversales al esqueleto principal. Rango de valores [0,100].	Color.
bidesv_r	Desviación estándar de los valores de iluminación en la máscara R del rango de colores RGB del objeto.	Color.

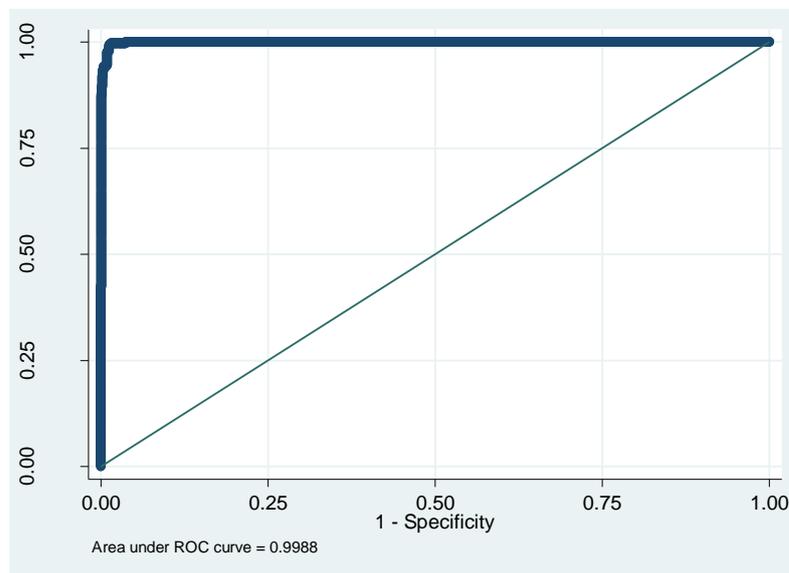
**Tabla1.** Variables involucradas en el modelo de objetos para el morfotipo *Lactobacillus*.

### *Análisis Estadístico del Modelo*

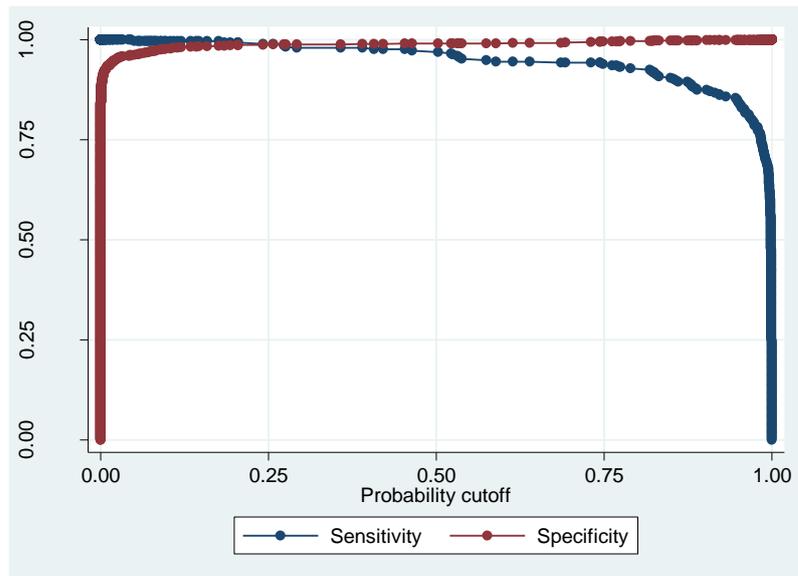
Las variables en la tabla son las que intervienen en el modelo logístico multivariado. Cada variable es significativas con un p-valor de 0.01 en el modelo univariado, salvo *Promfor* variable con un p-valor de 0.8 en el análisis univariado pero es significativa en el multivariado por tener una coherencia biológica es incluida en el

análisis. Además debemos mencionar que la variable con mejor poder de predicción individual es *prompuntasredesst* variable adimensional relacionada con la forma. Por la importancia de la elección de las variables independientes en la creación del modelo logístico múltiple se realizó un análisis de correlación múltiple, el máximo valor observado en él fue de 0,7181 adecuado para nuestro estudio.

El modelo logístico multivariado presento un Pseudo-R<sup>2</sup> de 93.05, en el intervinieron 8 variables con un máximo P-valor de 0.016. El área ROC para los resultados de nuestro modelo fue de 0,999 (Figura 10). El índice de Youden que maximiza el valor de la sensibilidad y especificidad presenta un valor de 0.967. Los valores de sensibilidad son de 97.63 y de especificidad de 99,10 con punto de corte en la probabilidad de 45.31, esto es, los objetos con una probabilidad del modelo mayor de 45.31 serán positivos y menor a ella negativos.



**Figura10.** Área ROC. Modelo de objeto Morfotipo *Lactobacillus*.



**Figura11.** Curva Lsens Modelo de objeto *Lactobacillus*. Sensibilidad Versus (1 - Especificidad). En la Figura 11 podemos observar la gráfica de comportamiento de la sensibilidad versus (1- especificidad) de nuestro diagnóstico, aquí se puede observar el alto valor predictivo de nuestro modelo estadístico, desde el comportamiento de la sensibilidad y especificidad en cada punto.

### 6.2.1.2. Modelo *Gardnerella*.

*Base de datos:*

La base de datos para *Gardnerella* considero una base conformada de 1331 objetos: 1016 *Gardnerella* y 315 objetos negativos conformados por *Lactobacillus*, *Mobiluncus* y restos celulares.

*Variables Significativas*

Se calcularon 184 características para cada objeto *Gardnerella* seleccionado. En la siguiente tabla se presenta la descripción de las variables significativas involucradas en el modelo estadístico. Entre las características resaltantes se encuentran tres de forma, de ellas una representa el comportamiento de los extremos del objeto

(*prompuntas\_re*), dos representan la variabilidad del espesor del objeto (*espesor\_redesvstan*, *std\_tam\_ram\_re*). La iluminación está presente por una variable (*lpromi\_re*); y el color de los objetos productos de la tinción Gram también por una variable (*spval*). Además solo una variable representa la importancia del tamaño en la elección de los objetos positivos para el modelo (*areabact*). Para mayor claridad ver Tabla.

Característica	Cuantificación	Representatividad
<i>areabact</i>	Área del objeto digital analizado en pixeles.	Tamaño
<i>prompuntas_re</i>	Promedio de las distancias entre los pixeles correspondientes opuestos a cada pixel extremo del esqueleto principal, relativo a la longitud del objeto.	Forma
<i>espesor_redesvstan</i>	Promedio de la desviación estándar de las longitudes de las líneas transversales al esqueleto principal separadas cada tres pixeles a lo largo de la longitud, normalizado por la longitud del objeto.	Forma
<i>std_tam_ram_re</i>	Desviación estándar de las medidas de las longitudes de los segmentos desde un punto extremo al esqueleto principal, relativas a la longitud del esqueleto principal.	Forma
<i>lpromi_re</i>	Promedio de los pixeles de la línea transversal al esqueleto principal del objeto, relativo a la longitud del objeto.	Iluminación
<i>spval</i>	Relación estructural entre el objeto analizado y el fondo de la fotografía.	Color

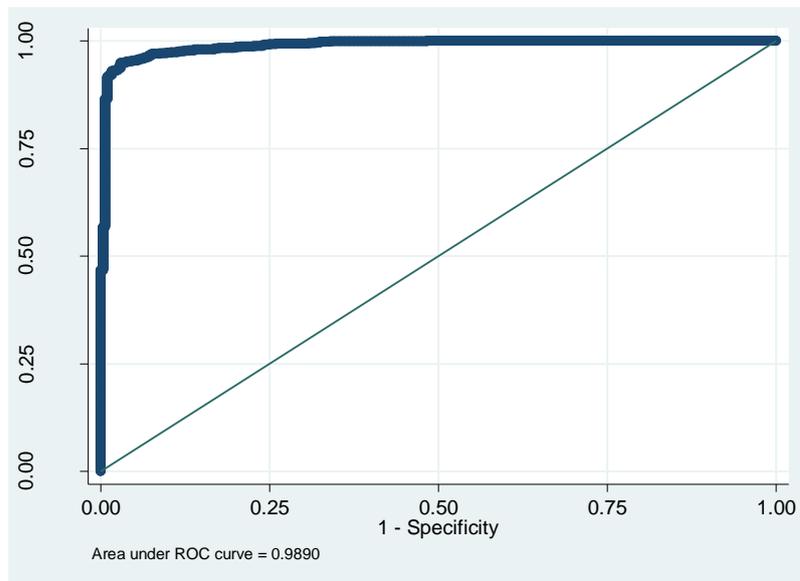
**Tabla2.** Variables involucradas en el modelo de objetos para el morfotipo *Gardnerellas*.

### *Análisis Estadístico del Modelo*

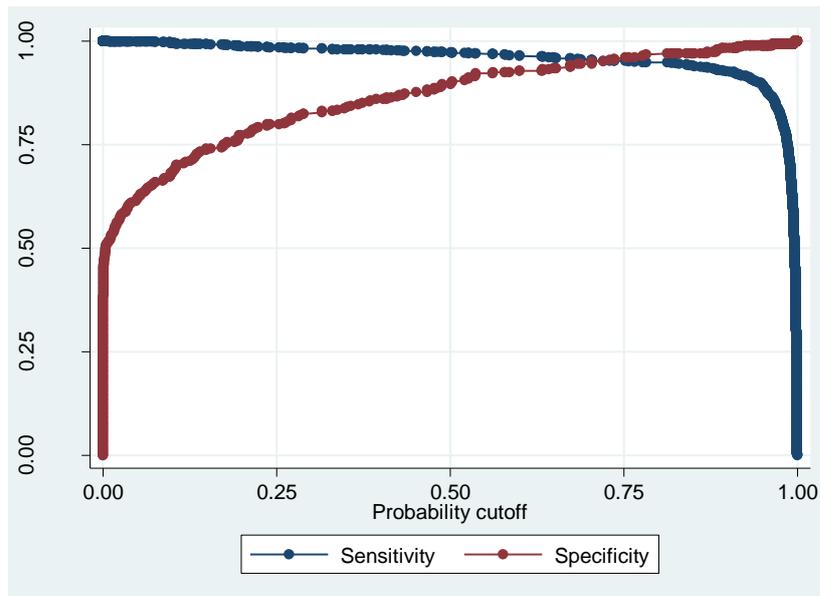
Las 6 variables en la tabla son las que intervienen en el modelo logístico multivariado. Cada variable es significativas con un p-valor de 0.001 en el modelo univariado. Además debemos mencionar que la variable con mejor poder de predicción individual es *areabact* variable dimensional relacionada con el tamaño del objeto. Por la importancia de la elección de las variables independientes en la

creación del modelo logístico múltiple se realizó un análisis de correlación múltiple, el máximo valor observado en él fue de 0,6572 adecuado para nuestro estudio.

El modelo logístico multivariado presento un Pseudo-R2 de 0.8020, en el intervinieron 6 variables con un máximo P-valor de 0.061. El área ROC para los resultados de nuestro modelo fue de 0,9890 (Figura 12). El índice de Youden que maximiza el valor de la sensibilidad y especificidad presenta un valor de 0.920. Los valores de sensibilidad son de 93.64 y de especificidad de 97.40 con punto de corte en la probabilidad de 81.25.



**Figura12.** Área ROC. Modelo de objeto Morfotipo *Gardnerellas*.



**Figura13.** Curva Lsens Modelo de objeto *Gardnerella*. Sensibilidad Versus (1 - Especificidad).

En la Figura 13 podemos observar la gráfica de comportamiento de la sensibilidad versus (1- especificidad) de nuestro diagnóstico, aquí se puede observar el decaimiento abrupto de la especificidad al aumentar la sensibilidad de nuestro modelo estadístico.

### 6.2.1.3. Modelo *Mobiluncus*.

*Base de datos:*

La base de datos para *Mobiluncus* considero una base conformada de 1331 observaciones: 1016 *Gardnerella* y 315 objetos negativos conformados por *Lactobacillus*, *Mobiluncus* y restos celulares.

*Variables Significativas*

Se calcularon 184 características para cada objeto *Mobiluncus* seleccionado. En la siguiente tabla se presenta la descripción de las variables significativas involucradas en el modelo estadístico. Entre las características resaltantes se encuentran una de forma que indica el comportamiento del espesor del objeto (*maxi\_espe\_re*). La

iluminación está presente por dos características una de la iluminación general del objeto (*bimedia*) y otra la variación de la iluminación en el interior del mismo (*lstdinti*); y el color de los objetos productos de la tinción Gram también por una variable (*bimedia\_r*).

Característica	Cuantificación	Representatividad
Perimetro	Número de pixeles del borde del objeto digital.	Tamaño.
maxi_espe_re	Razón de La longitud del máximo segmento de las líneas transversales al esqueleto principal del objeto analizado, con respecto a la longitud total del objeto	Forma.
bimedia	Promedio de los valores de iluminación en escala de grises del objeto.	Iluminación.
lstdinti	Desviación estándar de los pixeles de la línea transversal al esqueleto principal del objeto, eliminando 3 pixeles de cada extremo del segmento interior.	Iluminación.
bimedia_r	Promedio de los valores de iluminación en la máscara R del rango de colores RGB del objeto.	Color.
cval	Relación de contraste entre el objeto analizado y el fondo de la fotografía.	Similaridad.
spval	Relación estructural entre el objeto analizado y el fondo de la fotografía.	Similaridad.

**Tabla3.** Variables involucradas en el modelo de objetos para el morfotipo *Mobiluncus*.

También tenemos 2 variables que evalúan la similaridad del objeto con el fondo del campo(*cval,spval*). Además una variable representa la importancia del tamaño en la elección de los objetos positivos para el modelo (*perímetro*). Para explicación detallada ver Tabla3.

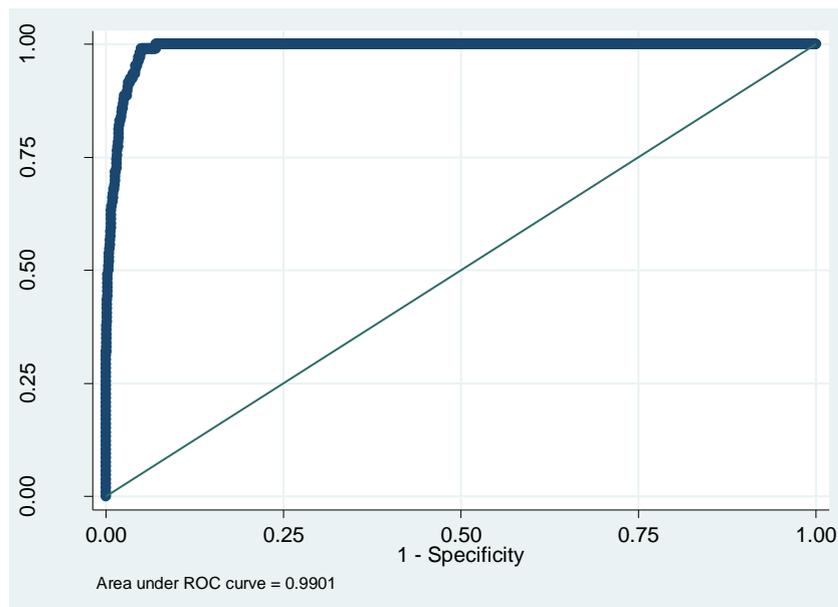
#### *Análisis Estadístico del Modelo*

Las variables en la tabla son las que intervienen en el modelo logístico multivariado. Cada variable es significativas con un p-valor menor de 0.004 en el modelo univariado. Además debemos mencionar que la variable con mejor poder de predicción individual es *bimedia* variable dimensional relacionada con la

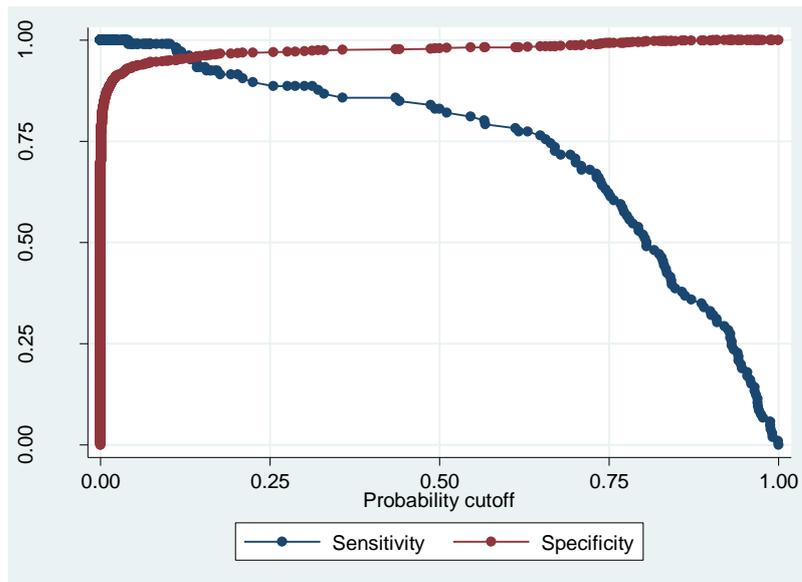
iluminación media del objeto. Por la importancia de la elección de las variables independientes en la creación del modelo logístico múltiple se realizó un análisis de correlación múltiple, el máximo valor observado en él fue de 0,7540.

El modelo logístico multivariado presento un Pseudo-R2 de 0.7507, en el intervinieron 7 variables con un máximo P-valor de 0.021. El área ROC para los resultados de nuestro modelo fue de 0,9901 (Figura 14). El índice de Youden que maximiza el valor de la sensibilidad y especificidad presenta un valor de 0.9170. Los valores de sensibilidad son de 96.23 y de especificidad de 95.51 con punto de corte en la probabilidad de 13.06.

En la Figura 15 podemos observar la gráfica de comportamiento de la sensibilidad versus (1- especificidad) de nuestro diagnóstico, aquí se puede observar el decaimiento abrupto de la sensibilidad al aumentar la especificidad de nuestro modelo estadístico.



**Figura14.** Área ROC. Modelo de objeto Morfotipo *Mobiluncus*.



**Figura15.** Curva Lsens Modelo de objeto *Mobiluncus*. Sensibilidad Versus (1 - Especificidad).

#### 6.2.1.4. Resumen de Modelos.

Los modelos de objetos de cada morfotipo son de vital importancia, en el diagnóstico de VB, presentamos un resumen de cada modelo seleccionado para cada morfotipo. En la tabla IV podemos observar los detalles numéricos mencionados con anticipación en esta sección, se puede visualizar además los cambios estadísticos de las características del análisis univariado y al análisis de regresión logística multivariado.

Los datos mas relevantes al calcular un modelo estadístico de identificación son resumidos en la siguiente tabla, donde observamos la sensibilidad, especificidad, el área bajo la curva ROC y el índice de Youden indicadores de la certeza y predicción de los modelos de clasificación para cada grupo de morfotipos involucrados en el diagnóstico de VB. Ver Tabla V.

Modelo	Pseudo R2	Características	Regresión Logística Univariada Odds Ratio (Pseudo R2, p-valor)	Regresión Logística Múltiple Odds Ratio ( p-valor )
<i>Lactobacillus</i>	0.9305	<i>Promptasredesst</i>	0.59 ( 0.4325 , <0.001 )	0.08 ( <0.001 )
		<i>Promptas</i>	0.88 (0.0046 , <0.01 )	47.10 ( <0.001 )
		<i>Espesordesvstand</i>	0.53 ( 0.0376, <0.001 )	0.004 ( <0.0001 )
		<i>Promfor</i>	0.99 ( 0.0001 , <0.8 )	0.60 ( <0.0001 )
		<i>bimedia</i>	0.96 (0.0892 , <0.001 )	0.79 ( <0.0001 )
		<i>bidesv_r</i>	1.29 ( 0.3377, <0.001 )	1.31 ( <0.0001 )
		<i>lpsval</i>	4.0E+05( 0.2421 , <0.001 )	2.01E-04 ( <0.02 )
<i>Gardnerella</i>	0.8020	<i>Spval</i>	6.61E-10 (0.2240 , <0.001 )	5.58E-11(<0.0001)
		<i>Espesor_redestand</i>	1.12( 0.272 , <0.001 )	1.15(<0.022)
		<i>Areabact</i>	0.98 (0.4589 , <0.0001)	0.99(<0.002)
		<i>Lpromi_re</i>	0.06 ( 0.1032 , <0.001 )	2.4E-03(<0.0001)
		<i>Promptas_re</i>	1.16 ( 0.3893 , <0.001 )	1.08(<0.0001)
		<i>Std_tam_ram_re</i>	2.48E-19 ( 0.158 , <0.001 )	4.37E-06(<0.062)
<i>Mobiluncus</i>	0.7507	<i>Cval</i>	0.013(0.0905, <0.0001)	1.03E-11(<0.0001)
		<i>Spval</i>	1.80 (0.001 , <0.0001)	4.15E-18(<0.0001)
		<i>maxi_espe_re</i>	0.89E-03(0.1437,<0.0001)	2.13E-15(<0.0001)
		<i>bimedia_r</i>	1.04(0.0113,<0.004)	0.81(<0.0001)
		<i>bimedia</i>	1.06(0.2061,<0.0001)	1.13(<0.0001)
		<i>lstdinti</i>	0.87 ( 0.0602 , <0.001 )	1.35(<0.022)
		<i>perimetro</i>	1.006(0.010,<0.003)	0.94(<0.0001)

**Tabla4.** Resumen de datos de los modelos de objetos por morfotipo. Incluye PseudoR2,- Odds ratio del modelo univariado y multivariado, y además P-valor.

Modelo De Objeto	Número de Observaciones	Pseudo R2	Sensibilidad	Especificidad	Area ROC	Índice Youden
<i>Lactobacillus</i>	1630	0.9305	97.63	99.10	0.999	0.967
<i>Gardnerella</i>	1331	0.8020	93.64	97.40	0.9890	0.920
<i>Mobiluncus</i>	1331	0.7507	96.23	95.51	0.9901	0.917

**Tabla5.** Resumen de resultados de los modelos de objetos por morfotipo. Incluye sensibilidad, especificidad, área ROC e índice de Youden.

### 6.2.1.5. Evaluación secuencial de modelos de objeto.

Se evaluó el modelo de objetos en una base de prueba con 1599 observaciones, escogidas en las condiciones no favorables. Los resultados de evaluación de dicha base que incluía una selección de *Lactobacillus*, *Gardnerellas* y *Mobiluncus*, brindo como resultado la siguiente tabla:

Modelo De Objeto	Número de Observaciones	Sensibilidad	Especificidad
<i>Lactobacillus</i>	1599	97.58	96.64
<i>Gardnerella</i>	1599	95.28	76.50
<i>Mobiluncus</i>	1599	94.55	89.12

**Tabla6.** Resumen de sensibilidad y especificidad de los modelos de objetos divididos por morfotipo. Incluye sensibilidad, especificidad en una base de datos de objetos de prueba.

Este análisis mostro resultados de menor especificidad para *Gardnerella* y *Mobiluncus* en condiciones reales, como vemos en la tabla anterior. Con el fin de mejorar los niveles de especificidad y sensibilidad en la identificación de morfotipos se eligió un modelo secuencial basado en la probabilidad de presencia de cada morfotipo involucrado.

La ausencia de *Lactobacillus* y la presencia de *Gardnerella* contribuyen en un 40%, cada una. Para determinar el orden inicial del análisis secuencial nos hemos basado en la incidencia de VB en población general.

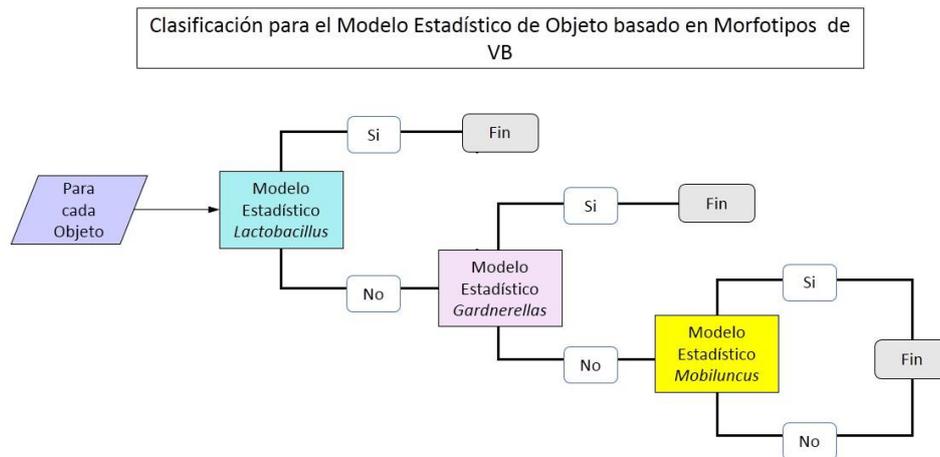
Estudios previos calculan una incidencia del 30% de VB en población general(10). Esto nos indica que aproximadamente un 70% de las mujeres poseen una flora normal, que se evidencia con la presencia de *Lactobacillus*.

Por ser mayor la población que posee estos morfotipos y su ausencia aumenta el score en el diagnóstico de VB, es elegido como el primer morfotipo del análisis secuencial.

El modelo para la identificación de *Gardnerellas* se coloca en segundo lugar del análisis debido a que contribuye con un mayor score al diagnóstico de VB, que la presencia de *Mobiluncus*.

De las publicaciones respecto al diagnóstico de VB, se conoce que la presencia de *Mobiluncus* aporta un máximo del 20% del diagnóstico y no está presente en todas las muestras ni los campos de las mismas, por lo que el modelo de identificación del mismo está ubicado en el tercer lugar del modelo secuencial.

Se consideró para el análisis de cada objeto en la fotografía, un análisis secuencial entre los modelos seleccionados como se ha explicado, así elevamos la especificidad del modelo. El análisis final para la identificación de los objetos en una foto se realiza con el diagrama secuencial visualizado en la imagen. Es así que cada objeto en la fotografía (1000 por foto aproximadamente) es evaluado por los modelos seleccionados de forma secuencial según el diagrama.



**Figura16.** Diagrama de evaluación secuencial del modelo de objeto en una fotografía.

Bajo el diagrama de análisis secuencial mostrado, se probó la especificidad de cada modelo en un grupo de prueba aumenta considerablemente. Especialmente se distingue un aumento de la especificidad en el modelo para la identificación de *Gardnerella*, los morfotipos más abundantes en las láminas con presencia de VB. Los resultados de este análisis se pueden visualizar en la siguiente tabla.

Modelo De Objeto	Número de Observaciones	Sensibilidad	Especificidad
<i>Lactobacillus</i>	1599	97.58	96.64
<i>Gardnerella</i>	1599	94.39	97.77
<i>Mobiluncus</i>	1599	74.55	97.85

**Tabla7.** Resumen de sensibilidad y especificidad de los modelos de objetos en un diseño de discriminación secuencial divididos por morfotipo. Incluye sensibilidad, especificidad en una base de datos de objetos de prueba.

Este modelo es el que se implementó en el software automático.

### 6.2.2. Modelo de fotos - Nivel 2.

El diagnóstico de VB para fotos es el segundo nivel de diagnóstico utilizado y se basa en el conteo y reconocimiento de los morfotipos involucrados.

Para el análisis estadístico de foto se extrajeron de cada fotografía 189 características.

Las características se pueden agrupar en:

Generales:

- Número de objetos en la foto.
- Iluminación de la fotografía.
- Iluminación del fondo.
- Iluminación de los objetos del campo.
- Medidas de similaridad: comparación de iluminación, estructura y contraste.

Modelos de Objetos

- 4 Modelos de objeto *Lactobacillus* – Número de Objetos Positivos
- 3 Modelos de objeto *Gardnerella* – Número de Objetos Positivos
- 3 Modelos de objeto *Mobiluncus* – Número de Objetos Positivos

Cabe mencionar que de estos modelos se ha seleccionado uno por morfotipo, el más relevante en los modelos estadísticos para fotos.

Las características son básicamente el número de objetos positivos para cada modelo de clasificación.

Para el modelo estadístico para fotos se han analizado de 1000 a 1500 objetos por foto aproximadamente. Cada foto se procesó aproximadamente en 35 segundos.

*Base de datos:*

Se crearon 2 bases de datos para fotos, la base de entrenamiento conformada por 50 fotos y la base de prueba conformada por 66 fotografías.

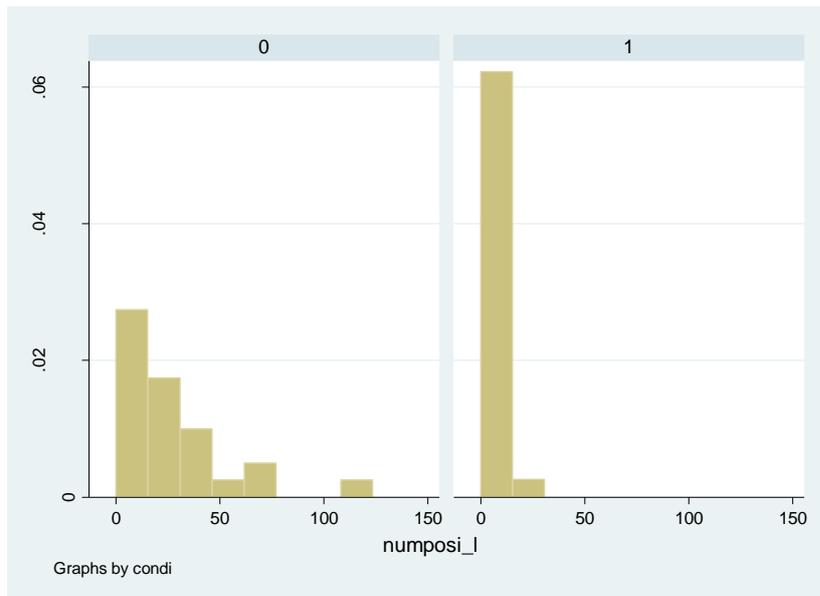
El Gold estándar en el modelo para fotos es el brindado por la Experta de la University of Washington la Dra. Karol Spiegel.

*Variables Significativas*

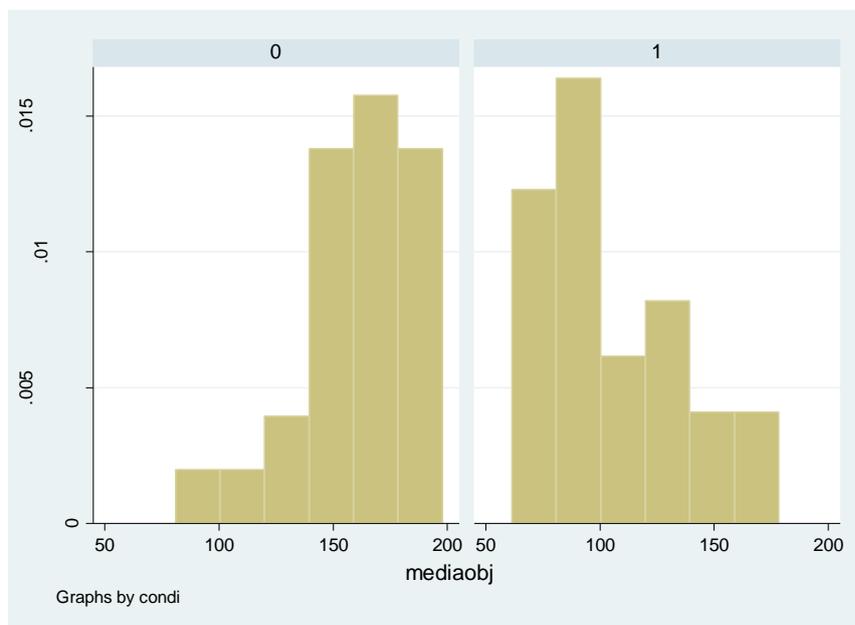
Se calcularon 189 características para cada fotografía de cada base. En la siguiente tabla se presenta la descripción de las 2 variables significativas involucradas en el modelo estadístico. Entre las características resaltantes se encuentran una directamente relacionada con el número de *Lactobacillus* (*numposi\_l*). La iluminación está presente por una características general basada en la iluminación media de los objetos en todo el campo (*mediaobj*).

Característica	Cuantificación	Representatividad
Numposi_l	Número de <i>Lactobacillus</i> presentes en la fotografía por el modelo de objeto Estadístico mencionado en la sección anterior.	<i>Lactobacillus</i> .
Mediaobj	Promedio de los valores de iluminación de todos los objetos segmentados en la fotografía.	Iluminación.

**Tabla8.** Variables involucradas en la formación del modelo de Fotos.



**Figura17.** Diagrama de evaluación secuencial del modelo de objeto en una fotografía.



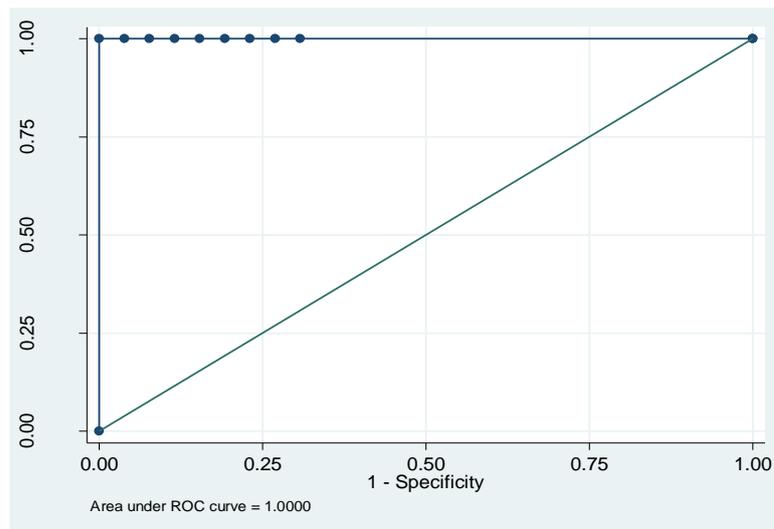
**Figura18.** Diagrama de evaluación secuencial del modelo de objeto en una fotografía.

### *Análisis Estadístico del Modelo*

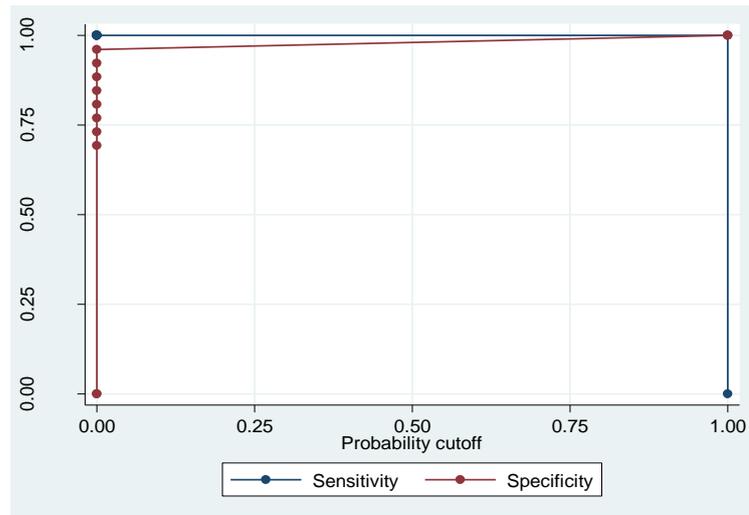
Las variables en la tabla son las que intervienen en el modelo logístico multivariado. Cada variable es significativa con un p-valor menor de 0.002 en el modelo univariado. Además debemos mencionar que la variable con mejor poder de

predicción individual es *numposi\_1* variable adimensional relacionada con la cantidad de *Lactobacillus* en la foto. Por la importancia de la elección de las variables independientes en la creación del modelo logístico múltiple se realizó un análisis de correlación múltiple, el máximo valor observado en él fue de 0,0388.

El modelo logístico multivariado presento un Pseudo-R2 de 1.00, en el intervinieron 2 variables. El área ROC para los resultados de nuestro modelo fue de 1.00 (Figura 19). El índice de Youden que maximiza el valor de la sensibilidad y especificidad presenta un valor de 1. Los valores de sensibilidad son de 100 y de especificidad de 96.15 con punto de corte en la probabilidad de 0.



**Figura19.** Área ROC. Modelo de Fotos.



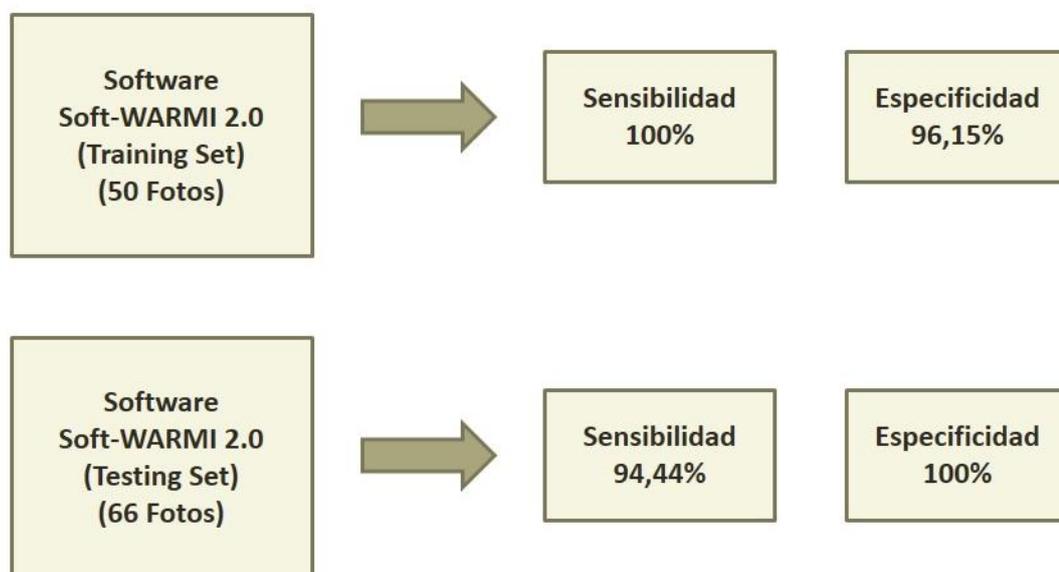
**Figura20.** Curva Lsens Modelo de Fotos. Sensibilidad Versus (1 - Especificidad).

En la Figura20 podemos observar la gráfica de comportamiento de la sensibilidad versus (1- especificidad) de nuestro diagnóstico.

Este modelo fue probado en la base de datos de prueba que conto con 66 fotografías, en este análisis se obtuvo una especificidad 100% y una sensibilidad 34.44%.

Los resultados se pueden visualizar en el siguiente cuadro resumen.

## Resultados – Modelo Foto



Lenguaje de Programación C.

Tiempo Aproximado de procesamiento por foto: 35 segundos.

**Figura21.** Resumen de resultados de modelo de fotos.

### *Evaluación de concordancia para el modelo de Foto.*

El análisis de concordancia de SoftWarmi con el Gold estándar (diagnóstico de Dr. Karol Spiegel) brindo como resultado un índice Kappa de 0.7769 y el porcentaje de Agreement 88.89%. Los resultados se pueden visualizar en la siguiente tabla.

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
88,89%	50,20%	0,7769	0,0923	8,41	0,0000

**Tabla9.** Resultados del análisis de concordancia para fotos(Gold Estándar - SoftWarmi).

Como referencia se analizó también la concordancia del Gold Estándar con el Diagnóstico de la Blga. Doris Lorena Lopez-Torres, entrenada en el Diagnóstico de VB en la University of Washington. La Blga. Doris Lorena contribuyo con el

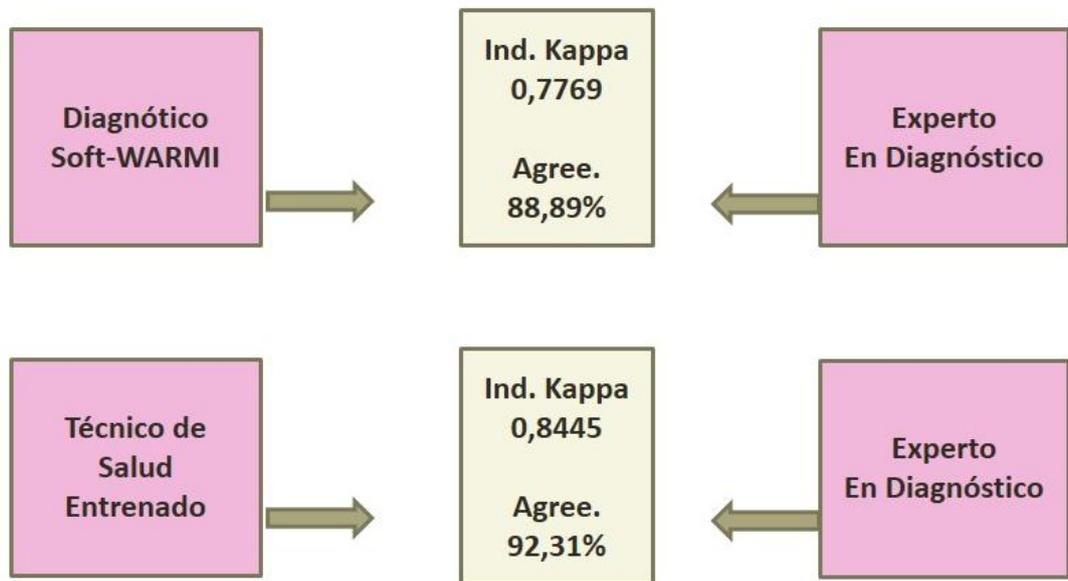
proyecto brindando el diagnóstico alterno de cada foto que conforma la base de prueba. Los resultados mostraron un porcentaje de agreement de 92.31% y un índice kappa de 0.8445, mostrados en la siguiente tabla.

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
92.31%	50,20%	0,8445	0,0924	9.14	0,0000

**Tabla10.** Resultados del análisis de concordancia para fotos(Gold Estándar – Biólogo Entrenado).

En el siguiente cuadro se observa un resumen de los niveles de concordancia entre SoftWarmi y la Blga. Lorena Torres.

## Resultados – Índice Kappa Foto



Gold Estándar: Dra. Karol Spiegel. University Of Washignton

**Figura22.** Resumen de resultados de concordancia para el modelo de foto.

### 6.2.3. Modelo de muestra - Nivel 3.

El diagnóstico de VB para muestras se realizó en base a las láminas de hisopados vaginales con tinción Gram proporcionadas por la University of Washington, bajo coordinaciones de la Dr. Patricia García, Decana de la Facultad de Salud Pública de la Universidad Peruana Cayetano Heredia.

#### *Base de datos:*

Se creó la base de datos para 59 muestras, las variables de cada muestra contaba con los diagnósticos de 10 fotografías por muestras. Las fotografías fueron capturadas por el Ingeniero Biotecnólogo Andrés Gutierrez Nuñez, quien fue entrenado por la Blga. Lorena Lopez-Torres, con el protocolo adecuado en la selección de campos válidos para el diagnóstico de VB.

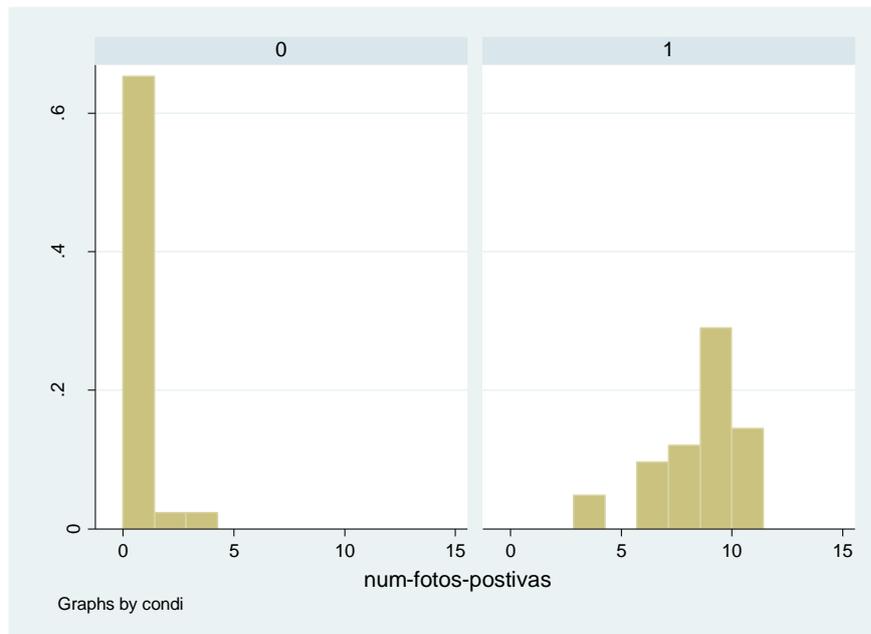
El Gold estándar en el modelo para muestras es el brindado por la University of Washington, pues las láminas son parte de material para educación en la institución.

#### *Variables Significativas*

Se calculó como variable la cantidad de campos positivos por muestra. En la siguiente tabla se presenta la descripción de la variable involucrada en el modelo estadístico.

Característica	Cuantificación	Representatividad
numfotospositivas	Número de fotografías por muestra con diagnóstico positivo para VB.	Campos positivos

**Tabla11.** Variable involucrada en el modelo de muestra.



**Figura23.** Distribución de Variable involucrada en el modelo de Muestra.

### *Resultados del Análisis Estadístico del Modelo*

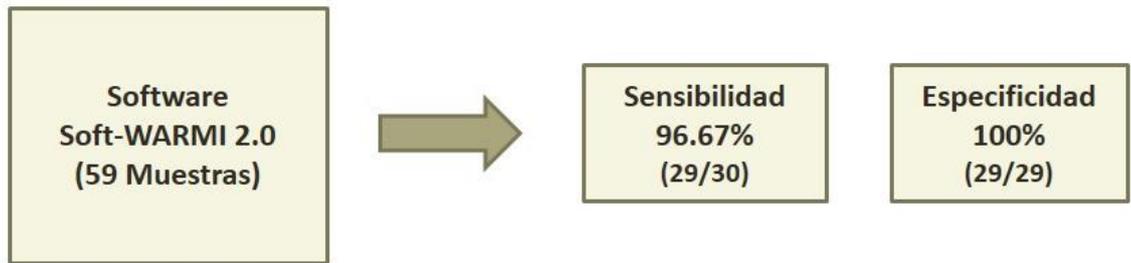
Para encontrar el número de fotos positivas que determinen la positividad de la muestra se realizó la regresión logística de la variable determinante. La regresión logística mostro como salida que con un número de fotos positivas  $>3$ , la base era perfectamente determinada.

Por este motivo se dicotomizó el diagnóstico del grupo de entrenamiento. Con la siguiente regla:

- Resul =1 ; Si numfotospositivas $\geq$ 3
- Resul =0 ; Otro caso

Esta regla se implementó en el grupo de entrenamiento de 60 muestras, en este análisis se obtuvo una especificidad 100%(29/29) y una sensibilidad 96.67%(29/30).

Los resultados se pueden visualizar en el siguiente cuadro resumen.



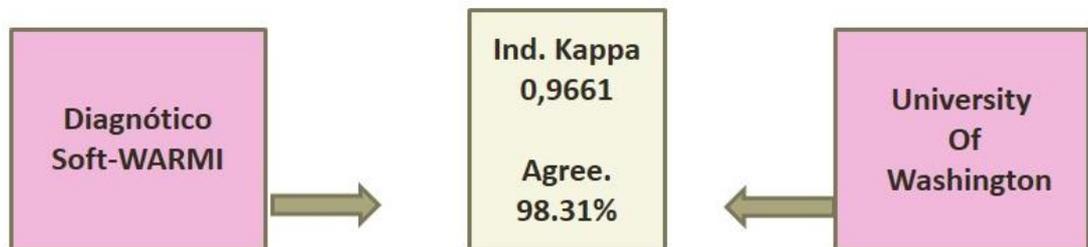
**Figura24.** Resumen del modelo de Muestra. Sensibilidad y Especificidad.

*Evaluación de concordancia para el modelo de Foto.*

El análisis de concordancia de SoftWarmi con el Gold estándar (diagnóstico University of Washington) brindo como resultado un índice Kappa de 0.9661 y el porcentaje de Agreement 98.31%. Los resultados se pueden visualizar en la siguiente tabla.

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
98.31%	49.99%	0.9661	0.1301	7.43	0,0000

**Tabla12.** Resultados del análisis de concordancia para diagnóstico de muestras (Gold Estándar – SoftWarmi).



**Figura25.** Resumen de análisis de concordancia en el diagnóstico por muestra para SoftWarmi.

## 7. DISCUSIÓN.

Durante la ejecución de este proyecto se ha demostrado que es posible realizar el diagnóstico automático de VB basado en imágenes digitales microscópicas de hisopados vaginales, con una sensibilidad del 96.67% y una especificidad de 100%. Para este resultado se desarrolló SoftWarmi, software automático basado en el reconocimiento de patrones, el software alcanzo una concordancia de 98.31% y un índice Kappa de 0.97 en el diagnóstico de muestras.

El diagnóstico basado en el procesamiento de imágenes, es una rama de investigación que ha crecido en los últimos años. Se han realizado algoritmos para cáncer de cuello uterino, malaria, entre otras. Pero, un software automático especializado en el diagnóstico de VB no se ha desarrollado hasta el momento, este trabajo es uno de los primeros para esta enfermedad en particular.

Durante los últimos años algunos estudios relacionados con el diagnóstico de vaginosis bacteriana se han enfocado en la necesidad de encontrar características y convertirlas en algoritmos de clasificación para predecir la presencia de la enfermedad. En el 2014 Baker (51) diseño algunos algoritmos de aprendizaje automático, este estudio conto con 418 variables entre las cuales, habían datos clínicos, macroscópicos, identificación de patógenos, y el olor, entre otros. Este estudio logro una clasificación con valores de exactitud sobre el 90% aproximadamente. En el 2015 Beck y Foster (52) realizaron diferentes variantes y uniones de los criterios de Amsel, Nugent y Ravel. Se crearon 3 modelos con 15 variables más importantes por cada uno y los consideran los mejores predictores de vaginosis bacteriana. Entre las características más importantes de estudio figuran

diferentes tipos de bacterias y morfotipos, además de datos clínicos como PH, olor, fluido vaginal, y algunos otros patógenos. Para clasificar la base de datos se utilizaron metodologías de aprendizaje automatizado como métodos aleatorios que utiliza arboles de decisión y programación genética, y además métodos estadísticos como regresión logística. Las tres opciones tuvieron resultados similares con una exactitud superior al 80%, siendo el más exacto el calculado a partir de la regresión logística. La regresión logística, método que también es utilizado en la clasificación en nuestra investigación. Es importante mencionar que de las 15 variables top para cada modelo elegido por los investigadores (53), la presencia de *Lactobacillus* y *Gardnerella* está presente en las variables de los tres modelos, lo que evidencia una presencia significativa en el análisis, así como en el desarrollo de nuestra investigación son variables significativas las también referidas a la presencia de *Gardnerella* y *Lactobacillus*. Los dos estudios mencionados si bien crean algoritmos para el diagnóstico de vaginosis bacteriana, estos se basan en datos clínicos, entre otros, la principal diferencia con SoftWarmi radica en el principio mismo de concepción, pues el algoritmo y diagnóstico que realizamos se basa únicamente en las fotografías capturadas de las muestras de hisopados vaginales con tinción gram, y en la extracción de características geométricas calculando un diagnóstico en base a la presencia o ausencia de los morfotipos involucrados en el diagnóstico de vaginosis bacteriana: *Lactobacillus*, *Gardnerella* y *Mobiluncus*.

Diferentes métodos de clasificación han sido utilizados para el reconocimiento automático de imágenes, entre ellos destacan las redes bayesianas, las redes neuronales (54), (55), SVM, MCSVM (56), sistemas neurodifusos (NEF, PROX, NEFCLASS, ANFIS) (57) y los modelos híbridos (58). Este estudio utiliza como

método de clasificación a la regresión logística logrando obtener resultados de sensibilidad y especificidad altas, comparables a otros estudios donde utilizan metodologías basadas en Machine Learning.

En el año 2013, Chen (59), se utilizó para el diagnóstico un dispositivo con sensores químicos para obtener los datos de entrada, y utiliza un procesamiento de señales para realizar un diagnóstico. Este dispositivo se utilizó para detectar cáncer en diferentes órganos, como promedio para este fin obtuvo una sensibilidad de 84.4% y una especificidad de 86.8%. En este mismo estudio se analiza el uso del dispositivo nariz electrónica, para la detección de infecciones en el tracto urinario por *Proteus spp*, *Staphylococcus* y *Escherichia coli*, pero no VB, ni asociados a *Gardnerella* con en el caso de SoftWarmi. En la investigación (59) se obtuvo un diagnóstico con una exactitud de 80% aproximadamente. La principal diferencia con nuestra propuesta radica en que este estudio recoge la información química de la muestra a través de un sistema basado en sensores metal-óxido, sensores conductores de polímero, sensores ópticos y sensores de microbalanza de cristal de cuarzo. En nuestra investigación no utilizamos sensores específicos, recolectamos las imágenes con microscopios y cámaras digitales acopladas y realizamos el análisis en base al diagnóstico de las imágenes de las muestras, con el análisis de sus características morfológicas, de color e iluminación. Utilizamos modelos estadísticos de regresión logística múltiple que nos permiten predecir la presencia de *Lactobacillus* y *Gardnerellas*, con lo cual concluimos un diagnóstico de VB. Otra diferencia sustancial entre estos estudios se basa en los métodos de clasificación nosotros utilizamos regresión logística múltiple y en estos estudios

utilizan el análisis de discriminante lineal y la metodología del análisis de las componentes principales.

Bengtsson y Malm en el año 2014 (34), publicaron una revisión bibliográfica de los métodos automáticos para el análisis de PAP, en el diagnóstico de cáncer cervical basado en el análisis de fotografías de muestras de Papanicolaou. Esta publicación analiza durante 3 décadas los avances en procesamiento de El reconocimiento automático de imágenes ha sido investigado en este estudios a partir de las características morfológicas del orden geométrico, de curvatura, de textura y color. Los algoritmos presentados en nuestro estudio consideran también estas características morfológicas como: longitud, espesor, color, variación de luminosidad interior, variación con la elipse de ajuste, etc; para cada uno de los morfotipos involucrados. Bengtsson y su equipo (34) analizan en este estudio la problemática en citopatología, como es la estandarización de la tinción en diferentes centros de salud, el problema de la superposición de la muestra, los contaminantes presentes, y las variaciones de iluminación que pueden afectar el diagnóstico por imágenes. Es importante resaltar que en el estudio de SoftWarmi, se realizaron protocolos de adquisición de fotografías y todas las láminas fueron teñidas por el mismo laboratorio mediante una técnica semejante, y las fotografías fueron adquiridas mediante similares condiciones por un solo sistema de adquisición de imágenes, para minimizar el error generado por las observaciones dadas en este estudio.

Nuestra investigación requiere fotografías de 10 campos de la muestra de hisopados vaginales. Nuestro Gold estándar es el score de Nugent la clasificación y el diagnostico se calcula por el conteo de morfotipos *Lactobacillus* y *Gardnerella*

identificadas por el reconocimiento de patrones y procesamiento de imágenes en la observación microscópica. Basándonos en estos principios la primera diferencia que encontramos con los métodos utilizados en las primeras publicaciones radica en que ellas están diseñadas sobre un conjunto de variables clínicas, de morfotipos presentes, e incluso otros patógenos, mientras que nuestro estudio solo en el procesamiento de imágenes. Además no utilizamos máquinas de aprendizaje, propiamente solo modelos estadísticos en base al conteo de los morfotipos en cada campo analizado.

Los recursos médicos apropiados para un diagnóstico adecuado de vaginosis bacteriana son escasos en todo el Perú. En las zonas altoandinas la presencia de especialistas de cualquier especialidad aun es menor, se encontró que en el 2010 solo estaban registrados en el minsa aproximadamente 70 patólogos a nivel nacional y este escenario no ha variado mucho a lo largo de estos años, sigue siendo una preocupante realidad.

El diagnóstico de VB es muy complicado y muestra una alta variabilidad inter e intra observador. Esto es consecuencia de la falta de homogeneidad de los distintos campos microscópicos, a la similitud entre algunos microorganismos, y a la necesidad de realizar el conteo de microorganismos que en cada campo a 1000x son abundantes.

SoftWarmi es una solución apropiada para mejorar la reproducibilidad de las lecturas, y de los diagnósticos, además puede ser usado como método de screening para reducir los efectos de la vaginosis bacteriana y en apoyo general a la salud femenina a nivel nacional.

Cabe mencionar que SoftWarmi va a permitir realizar estudios descriptivos poblacionales, y evaluar niveles de prevalencia, incidencia, etc.

SoftWarmi puede usarse como una herramienta diagnóstica a bajo costo, accesible, disponible y con una cualidad de reproducibilidad, basada en el método de diagnóstico de Nugent para vaginosis bacteriana.

Con los avances realizados nuestro siguiente paso es trabajar por la implementación del telediagnóstico de vaginosis bacteriana en lugares remotos donde no existen expertos y donde los recursos son insuficientes para establecer un sistema de salud viable y orientada a la salud femenina.

Podemos concluir que durante este estudio se ha desarrollado una herramienta útil para facilitar el diagnóstico de vaginosis bacteriana de manera masiva, que se puede acoplar a sistemas remotos para hacerla accesible a poblaciones con escasos recursos médicos en bien de la población peruana.

## **8. LIMITACIONES.**

El estudio tiene una importante limitación:

- Todos nuestros resultados dependen de la lectura y selección de campos realizada por el tecnólogo médico encargado. Esto puede verse alterado por la sensibilidad y especificidad del técnico de salud para la selección del campo idóneo para el diagnóstico. Cabe mencionar que el profesional ha sido capacitado para este fin anteriormente de su participación en este estudio.

## 9. CONCLUSIONES Y RECOMENDACIONES.

Durante la ejecución de este estudio se ha podido concluir de manera global que que ha sido posible desarrollar un software de diagnóstico automático de vaginosis bacteriana basado en el reconocimiento y conteo de los morfotipos involucrados como son *Lactobacillus*, *Gardnerella* y *Mobiluncus*.

Una de los hallazgos más importantes en el análisis es que el conteo de *Mobiluncus* no resulto significativo para la determinación de VB en fotos. El número de *Mobiluncus* obtuvo en promedio de los 3 modelos de clasificación obtenidos un P-valor de 0.0623, y valores de P-seudo R2 de 12.45, que explicaría solo en el 12% de los casos del evento y no resulta significativo en el análisis. A pesar de este detalle se incluyó en los modelos de clasificación y no apporto significativamente en el análisis.

El diagnóstico de VB en fotos, se calculó en base al conteo de los *Lactobacillus* en las fotografías de las muestras y en la iluminación media de los objetos. Con esto podemos inferir que es necesario y suficiente para nuestro algoritmo el reconocimiento de los *Lactobacillus* y el conteo de los mismos para determinar un diagnóstico positivo de VB.

Cabe mencionar que se encontraron otros modelos de fotografía que implicaban la importancia del conteo de *Lactobacillus* y *Gardnerellas*, pero estos no presentaron tan buenos resultados como el modelo explicado y detallado en el documento.

El porcentaje de Agreement de fotos que compara el diagnóstico de SoftWarmi 88.89% con la técnico de salud entrenada 92.31%, calcula un 3.4% de diferencias entre los diagnósticos de más de 100 fotos. Con esto podemos concluir que

SoftWarmi logro brindar resultados aceptables, comparables al técnico entrenado en la University of Washington.

Además es importante mencionar que el porcentaje de agreement para el diagnóstico de muestras de 97.62% nos permite concluir es posible implementar un sistema de telediagnóstico de VB con la herramienta desarrollado en favor de la salud sexual femenina en el Perú. Además se pueden realizar estudios descriptivos poblacionales para determinar la prevalencia de VB en determinadas poblaciones con características específicas.

Entre algunos de los siguientes pasos siguientes podemos pensar incluso en la posibilidad de generar nuevos métodos de diagnósticos basado en el reconocimiento de morfotipos o evaluar la significancia de cada morfotipo en el diagnóstico de VB, con estudios en poblaciones específicas.

## 10. BIBLIOGRAFÍA.

1. García PJ, Cárcamo CP, Chiappe M, Holmes KK. Sexually transmitted and reproductive tract infections in symptomatic clients of pharmacies in Lima, Peru. *Sex Transm Infect.* 2007 Apr;83(2):142–6.
2. Cammack AL, Buss C, Entringer S, Hogue CJ, Hobel CJ, Wadhwa PD. The association between early life adversity and bacterial vaginosis during pregnancy. *Am J Obstet Gynecol.* 2011 May;204(5):431.e1–8.
3. McCoy MC, Katz VL, Kuller JA, Killam AP, Livengood CH 3rd. Bacterial vaginosis in pregnancy: an approach for the 1990s. *Obstet Gynecol Surv.* 1995 Jun;50(6):482–8.
4. McGregor JA, French JI, Seo K. Premature rupture of membranes and bacterial vaginosis. *Am J Obstet Gynecol.* 1993 Aug;169(2 Pt 2):463–6.
5. Goldenberg RL, Andrews WW, Yuan AC, MacKay HT, St Louis ME. Sexually transmitted diseases and adverse outcomes of pregnancy. *Clin Perinatol.* 1997 Mar;24(1):23–41.
6. Simhan HN, Bodnar LM, Krohn MA. Paternal race and bacterial vaginosis during the first trimester of pregnancy. *Am J Obstet Gynecol.* 2008 Feb;198(2):196.e1–4.
7. Cohen CR, Lingappa JR, Baeten JM, Ngayo MO, Spiegel CA, Hong T, et al. Bacterial vaginosis associated with increased risk of female-to-male HIV-1 transmission: a prospective cohort analysis among African couples. *PLoS Med.* 2012;9(6):e1001251.
8. Africa CWJ. Efficacy of methods used for the diagnosis of bacterial vaginosis. *Expert Opin Med Diagn.* 2013 Mar;7(2):189–200.

9. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol.* 1991 Feb;29(2):297–301.
10. Jones FR, Miller G, Gadea N, Meza R, Leon S, Perez J, et al. Prevalence of bacterial vaginosis among young women in low-income populations of coastal Peru. *Int J STD AIDS.* 2007 Mar;18(3):188–92.
11. Sánchez SE, Koutsky LA, Sánchez J, Fernández A, Casquero J, Kreiss J, et al. Rapid and inexpensive approaches to managing abnormal vaginal discharge or lower abdominal pain: an evaluation in women attending gynaecology and family planning clinics in Peru. *Sex Transm Infect.* 1998 Jun;74 Suppl 1:S85–94.
12. Medina R, Rechkemmer A, Garcia-Hjarles M. Prevalencia de vaginitis y vaginosis bacteriana en pacientes con flujo vaginal anormal en el Hospital Nacional Arzobispo Loayza. *Rev Medica Hered.* 1999 Oct;10(4):144–50.
13. Spiegel CA. New developments in the etiology and pathogenesis of bacterial vaginosis. *Adv Exp Med Biol.* 1987;224:127–34.
14. Gillet E, Meys JFA, Verstraelen H, Verhelst R, De Sutter P, Temmerman M, et al. Association between bacterial vaginosis and cervical intraepithelial neoplasia: systematic review and meta-analysis. *PloS One.* 2012;7(10):e45201.
15. García PJ, Chavez S, Feringa B, Chiappe M, Li W, Jansen KU, et al. Reproductive tract infections in rural women from the highlands, jungle, and coastal regions of Peru. *Bull World Health Organ.* 2004 Jul;82(7):483–92.
16. Biswas MK. Bacterial vaginosis. *Clin Obstet Gynecol.* 1993 Mar;36(1):166–76.

17. Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am J Med.* 1983 Jan;74(1):14–22.
18. Notelovitz M, Massel P, Jones W. Asymptomatic vaginal disease in pregnancy. *South Afr Med J Suid-Afr Tydskr Vir Geneesk.* 1974 Apr 17;48(18):803–6.
19. Sobel JD. Vaginal infections in adult women. *Med Clin North Am.* 1990 Nov;74(6):1573–602.
20. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol.* 1991 Feb;29(2):297–301.
21. Spiegel CA, Amsel R, Holmes KK. Diagnosis of bacterial vaginosis by direct gram stain of vaginal fluid. *J Clin Microbiol.* 1983 Jul;18(1):170–7.
22. Redelinghuys MJ, Ehlers MM, Dreyer AW, Kock MM. Normal flora and bacterial vaginosis in pregnancy: an overview. *Crit Rev Microbiol.* 2015 Apr 2;1–12.
23. Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, et al. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. *N Engl J Med.* 1995 Dec 28;333(26):1737–42.
24. Gibbs RS. Chorioamnionitis and bacterial vaginosis. *Am J Obstet Gynecol.* 1993 Aug;169(2 Pt 2):460–2.
25. Ugwumadu A. Chorioamnionitis and mid-trimester pregnancy loss. *Gynecol Obstet Invest.* 2010;70(4):281–5.

26. Hillier SL, Kiviat NB, Hawes SE, Hasselquist MB, Hanssen PW, Eschenbach DA, et al. Role of bacterial vaginosis-associated microorganisms in endometritis. *Am J Obstet Gynecol.* 1996 Aug;175(2):435–41.
27. Faro S, Martens M, Maccato M, Hammill H, Pearlman M. Vaginal flora and pelvic inflammatory disease. *Am J Obstet Gynecol.* 1993 Aug;169(2 Pt 2):470–4.
28. Sweet RL. Role of bacterial vaginosis in pelvic inflammatory disease. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 1995 Jun;20 Suppl 2:S271–5.
29. Soper DE. Bacterial vaginosis and postoperative infections. *Am J Obstet Gynecol.* 1993 Aug;169(2 Pt 2):467–9.
30. van Oostrum N, De Sutter P, Meys J, Verstraelen H. Risks associated with bacterial vaginosis in infertility patients: a systematic review and meta-analysis. *Hum Reprod Oxf Engl.* 2013 Mar 29;
31. Koumans EH, Sternberg M, Bruce C, McQuillan G, Kendrick J, Sutton M, et al. The prevalence of bacterial vaginosis in the United States, 2001-2004; associations with symptoms, sexual behaviors, and reproductive health. *Sex Transm Dis.* 2007 Nov;34(11):864–9.
32. Schmid GP. The epidemiology of bacterial vaginosis. *Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet.* 1999 Nov;67 Suppl 1:S17–20.
33. Guédou FA, Van Damme L, Mirembe F, Solomon S, Becker M, Deese J, et al. Intermediate vaginal flora is associated with HIV prevalence as strongly as bacterial vaginosis in a cross-sectional study of participants screened for a randomised controlled trial. *Sex Transm Infect.* 2012 Nov;88(7):545–51.

34. Bengtsson E, Malm P. Screening for cervical cancer using automated analysis of PAP-smears. *Comput Math Methods Med.* 2014;2014:842037.
35. Lu H, Jiang P-C, Zhang X-D, Hou W-J, Wei Z-H, Lu J-Q, et al. Characteristics of bacterial vaginosis infection in cervical lesions with high risk human papillomavirus infection. *Int J Clin Exp Med.* 2015;8(11):21080–8.
36. Salah RM, Allam AM, Magdy AM, Mohamed AS. Bacterial vaginosis and infertility: cause or association? *Eur J Obstet Gynecol Reprod Biol.* 2013 Mar;167(1):59–63.
37. Denney JM, Culhane JF. Bacterial vaginosis: a problematic infection from both a perinatal and neonatal perspective. *Semin Fetal Neonatal Med.* 2009 Aug;14(4):200–3.
38. Sexually Transmitted Diseases Treatment Guidelines, 2006 [Internet]. Medscape. [cited 2016 Apr 21]. Available from: <http://www.medscape.com/viewarticle/543426>
39. Donders GGG, Zozzika J, Rezeberga D. Treatment of bacterial vaginosis: what we have and what we miss. *Expert Opin Pharmacother.* 2014 Apr;15(5):645–57.
40. Foran DJ, Comaniciu D, Meer P, Goodell LA. Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. *IEEE Trans Inf Technol Biomed Publ IEEE Eng Med Biol Soc.* 2000 Dec;4(4):265–73.
41. Sampat MP, Bovik AC, Aggarwal JK, C KRC. Supervised Parametric and Non-parametric Classification of Chromosome Images. *Pattern Recognit.* 2005;38:1209–23.
42. Bottigli U, Carpinelli M, Fiori PL, Golosio B, Marras A, Masala GL, et al. A New Automatic System of Cell Colony Counting.

43. Trattner S, Greenspan H, Abboud S. Automatic Identification of Bacterial Types Using Statistical Imaging Methods. In: Proceedings of SPIE International Symposium on Medical Imaging. 2003.
44. Ross NE, Pritchard CJ, Rubin DM, Dusé AG. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Med Biol Eng Comput.* 2006 May;44(5):427–36.
45. Castanon C, Fraga J, Fernandez S, Gruber A, Dafcosta L. Biological shape characterization for automatic image recognition and diagnosis of protozoan parasites of the genus *Eimeria*☆. *Pattern Recognit.* 2007 Jul;40(7):1899–910.
46. Alva A, Aquino F, Gilman RH, Olivares C, Requena D, Gutiérrez AH, et al. Morphological Characterization of *Mycobacterium tuberculosis* in a MODS Culture for an Automatic Diagnostics through Pattern Recognition. *PLoS ONE.* 2013 Dec 16;8(12):e82809.
47. Fitzgibbon A, Pilu M, Fisher RB. Direct least square fitting of ellipses. *IEEE Trans Pattern Anal Mach Intell.* 1999 May;21(5):476–80.
48. Flusser J. On the inverse problem of rotation moment invariants. *Pattern Recognit.* 2002 Dec;35(12):3015–7.
49. Dutt A, Rokhlin V. Fast Fourier Transforms for Nonequispaced Data. *SIAM J Sci Comput.* 1993 Nov 1;14(6):1368–93.
50. Becher PF, Sun EY, Plucknett KP, Alexander KB, Hsueh C-H, Lin H-T, et al. Microstructural Design of Silicon Nitride with Improved Fracture Toughness: I, Effects of Grain Shape and Size. *J Am Ceram Soc.* 1998 Nov 1;81(11):2821–30.

51. Baker YS, Agrawal R, Foster JA, Beck D, Dozier G. APPLYING MACHINE LEARNING TECHNIQUES IN DETECTING BACTERIAL VAGINOSIS. Proc Int Conf Mach Learn Cybern Int Conf Mach Learn Cybern. 2014 Jul 3;2014:241–6.
52. Beck D, Foster JA. Machine learning classifiers provide insight into the relationship between microbial communities and bacterial vaginosis. BioData Min. 2015;8:23.
53. Beck D, Foster JA. Machine learning techniques accurately classify microbial communities by bacterial vaginosis characteristics. PloS One. 2014;9(2):e87830.
54. Yang YS, Park DK, Kim HC, Choi M-H, Chai J-Y. Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. IEEE Trans Biomed Eng. 2001 Jun;48(6):718–30.
55. A Sengur, I Turkoglu. Parasite egg cell classification using invariant moments. Proc 4th Int Symp Intell Manuf Syst. 2004;98–106.
56. Avci D, Varol A. An expert diagnosis system for classification of human parasite eggs based on multi-class SVM. Expert Syst Appl. 2009 Jan;36(1):43–8.
57. Dogantekin E, Yilmaz M, Dogantekin A, Avci E, Sengur A. A robust technique based on invariant moments – ANFIS for recognition of human parasite eggs in microscopic images. Expert Syst Appl. 2008 Oct;35(3):728–38.
58. Ghazali KH, Alsameraai RSH, Mohamed Z. Automated System for Diagnosis Intestinal Parasites by Computerized Image Analysis. Mod Appl Sci [Internet]. 2013 Apr 25 [cited 2016 Apr 22];7(5). Available from: <http://www.ccsenet.org/journal/index.php/mas/article/view/23609>
59. Chen S, Wang Y, Choi S. Applications and Technology of Electronic Nose for Clinical Diagnosis. Open J Appl Biosens. 2013;02(02):39–50.