



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

**DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA
INFORMACIÓN**

**MODELO GRÁFICO CAUSAL PARA REPRESENTAR LAS
BACTERIAS COEXISTENTES DETONANTES DE VAGINOSIS
BACTERIANA EN MUJERES EMBARAZADAS**

**TESIS PARA OBTENER EL GRADO DE:
DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN**

PRESENTA:

MARICELA GARCÍA AVALOS

BAJO LA DIRECCIÓN DE:

DRA. JUANA CANUL REICH

EN CODIRECCIÓN:

DRA. LIL MARÍA XIBAI RODRÍGUEZ HENRÍQUEZ

CUNDUACÁN, TABASCO, A: JUNIO 2026



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

**DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA
INFORMACIÓN**

**MODELO GRÁFICO CAUSAL PARA REPRESENTAR LAS
BACTERIAS COEXISTENTES DETONANTES DE VAGINOSIS
BACTERIANA EN MUJERES EMBARAZADAS**

**TESIS PARA OBTENER EL GRADO DE:
DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN**

PRESENTA:

MARICELA GARCÍA AVALOS

BAJO LA DIRECCIÓN DE:

DRA. JUANA CANUL REICH

EN CODIRECCIÓN:

DRA. LIL MARÍA XIBAI RODRÍGUEZ HENRÍQUEZ

CUNDUACÁN, TABASCO, A: JUNIO 2026

Declaración de Autoría y Originalidad

En la Ciudad de Cunduacán el día Veinticuatro del mes de Junio del año 2026, el que suscribe **Maricela García Avalos**, alumna del Programa de la **Doctorado en Ciencias de la Computación** con número de matrícula **231H18006**, adscrito a la **División Académica de Ciencias y Tecnologías de la Información**, de la Universidad Juárez Autónoma de Tabasco, como autor de la Tesis presentada para la obtención de Grado de Doctorado y titulada **Modelo gráfico causal para representar las bacterias coexistentes detonantes de vaginosis bacteriana en mujeres embarazadas**, dirigida por el Dra. Juana Canul Reich y la Dra. Lil María Xibai Rodríguez Henríquez.

DECLARO QUE: La Tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la LEY FEDERAL DEL DERECHO DE AUTOR (Decreto por el que se reforman y adicionan diversas disposiciones de la Ley Federal del Derecho de Autor del 01 de Julio de 2020 regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita. Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad o contenido de la Tesis presentada de conformidad con el ordenamiento jurídico vigente.

Cunduacán, Tabasco a 24 de Junio de 2026.



Estudiante: Maricela García Avalos



UJAT
UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



2026
año de
Margarita
Maza

Cunduacán, Tabasco, a 24 de junio de 2026
Oficio No. 1321/2026/DACYTI/D

Asunto: Autorización de impresión de Tesis

C. Maricela García Avalos

Egresada del Doctorado en Ciencias de la Computación

En virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudios de Posgrado vigente en la Universidad, informo a Usted que se autoriza la impresión del trabajo recepcional **"Modelo gráfico causal para representar las bacterias coexistentes detonantes de vaginosis bacteriana en mujeres embarazadas"**, para presentar examen y obtener el Grado de Doctora en Ciencias de la Computación.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

Atentamente


Dra. Laura Beatriz Vidal Turrubiates
Directora



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS

C.c.p. Dr. Eduardo Hernández de la Cuz. - Encargado del despacho de la Coordinación de Posgrado.
Alumno.

DRA.*LBVT/EBG

Av. Universidad s/n, Zona de la Cultura, Col. Magisterial,
Villahermosa, Centro, Tabasco, Mex. C.P. 86040.
Tel (993) 358 15 00 e-Mail: rectoria@ujat.mx

Carta de Cesión de Derechos

Villahermosa, Tabasco a 24 de Junio de 2026.

Por medio de la presente manifiesto haber colaborado como AUTOR en la producción, creación y/o realización de la obra denominada: **Modelo gráfico causal para representar las bacterias coexistentes detonantes de vaginosis bacteriana en mujeres embarazadas** .

Con fundamento en el artículo 83 de la Ley Federal del Derecho de Autor y toda vez que, la creación y/o realización de la obra antes mencionada se realizó bajo la comisión de la Universidad Juárez Autónoma de Tabasco; entendemos y aceptamos el alcance del artículo en mención de que tenemos el derecho al reconocimiento como autores de la obra, y a la Universidad Juárez Autónoma de Tabasco mantendrá en un 100% la titularidad de los derechos patrimoniales por un período de 20 años sobre la obra en la que colaboramos, por lo anterior, cedemos el derecho patrimonial exclusivo en favor de la Universidad.

COLABORADOR



Estudiante: Maricela García Avalos

TESTIGOS



Dra. Juana Canul Reich

LIL MARIA R.

Dra. Lil María Xibai Rodríguez Henríquez

Dedicatoria

A mis padres, por su amor incondicional, su fortaleza y sus enseñanzas, que han sido una guía constante en mi vida.

A mis hijos, quienes son mi mayor inspiración y la motivación que me impulsa a seguir adelante cada día.

A mi esposo, por su apoyo, comprensión y compañía a lo largo de cada etapa de este camino.

Este logro también les pertenece, porque su cariño, paciencia y confianza hicieron posible la culminación de esta meta doctoral.

Universidad Juárez Autónoma de Tabasco.
México.

Agradecimientos

A Dios, por ser mi guía y fortaleza en cada momento de este proceso doctoral. Gracias por darme salud, paciencia y la oportunidad de alcanzar una meta más en mi vida académica y personal.

A mi familia, por estar siempre presente con su cariño, paciencia y apoyo incondicional. Gracias por creer en mí y ser el motor que me impulsó a culminar esta importante etapa de mi vida.

A mis amigos, por su amistad, apoyo y comprensión durante este proceso doctoral. Su compañía y palabras de ánimo fueron importantes para alcanzar esta meta.

A mis directoras de tesis, Dra. Juana Canul Reich y Dra. Lil María Xibai Rodríguez Henríquez, por su orientación académica, apoyo y valiosas aportaciones durante el desarrollo de este trabajo doctoral.

Al Dr. Erick Natividad de la Cruz Hernández, por facilitar el conjunto de datos utilizado en esta investigación y por sus aportaciones en la evaluación biológica de los resultados.

A los profesores del programa doctoral, por sus enseñanzas, orientación y valiosas aportaciones durante mi formación académica, las cuales contribuyeron al desarrollo de esta investigación.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), por el apoyo económico otorgado mediante la beca de posgrado número 842958 para la realización de mis estudios doctorales.

Al Sindicato de Profesores e Investigadores de la Universidad Juárez Autónoma de Tabasco (SPIUJAT), por el apoyo y las gestiones administrativas brindadas durante este proceso doctoral.

A la División Académica de Ciencias y Tecnologías de la Información (DACyTI) y a la Universidad Juárez Autónoma de Tabasco (UJAT), por el respaldo académico e institucional otorgado durante mi formación doctoral.

Índice general

Índice de Tablas	V
Índice de Figuras	VII
Resumen	IX
Abstract	X
1. Protocolo de tesis	1
1.1. Introducción	1
1.2. Marco teórico	4
1.2.1. Inteligencia artificial	4
1.2.2. Aprendizaje automático	4
1.2.2.1. Aprendizaje supervisado	4
1.2.2.2. Aprendizaje no supervisado	5
1.2.3. Big data	5
1.2.4. Lenguaje de programación R	5
1.2.5. Vaginosis bacteriana (VB)	5
1.2.5.1. Vaginosis bacteriana en embarazadas	6
1.2.6. Modelos de ecuaciones estructurales (MEE)	6
1.2.6.1. Tipos de modelos de ecuaciones estructurales	6
1.3. Justificación	7
1.4. Pregunta de investigación	7
1.5. Hipótesis	8

1.6. Objetivo general	8
1.6.1. Objetivos específicos	8
1.7. Metodología	8
1.8. Cronograma de actividades	10
2. Hacia un modelo gráfico causal para el análisis de la vaginosis bacteriana	13
2.1. Introducción	15
2.2. Materiales y métodos	16
2.2.1. Matriz de correlación	16
2.2.2. Modelo de ecuaciones estructurales	17
2.2.3. Métricas estadísticas de MEE	17
2.3. Estudio experimental	18
2.3.1. Conjunto de datos VB	19
2.3.2. Estructura del conjunto de datos	19
2.3.3. Eliminación de atributos	19
2.3.4. Valores atípicos	20
2.3.5. Valores faltantes	20
2.3.6. Matriz de correlación de Pearson	22
2.3.7. Modelo gráfico causal	24
2.3.8. Validación biológica del MGC	26
2.4. Resultados	27
2.5. Conclusiones	27
2.6. Declaración de contribución de autores y colaboradores	28
3. Modelo gráfico causal de la vaginosis bacteriana en mujeres embarazadas	31
3.1. Introducción	33
3.2. Materiales y métodos	34
3.2.1. Matriz de correlación de Spearman	34
3.2.2. Análisis de trayectorias	35
3.2.3. Métricas estadísticas en análisis de trayectorias	36
3.2.4. Prueba de Kolmogorov-Smirnov	37

3.2.5. Gráficos Q-Q	37
3.2.6. Transformación de datos	38
3.3. Estudio experimental	39
3.3.1. Preprocesamiento de datos	39
3.3.2. Aplicación del análisis de trayectorias	43
3.4. Resultados	51
3.5. Discusión	52
3.6. Conclusiones	52
4. Modelos gráficos causales para la transición de una microbiota vaginal saludable a la vaginosis bacteriana en mujeres embarazadas	57
4.1. Introducción	59
4.2. Materiales y métodos	60
4.2.1. Matriz de correlación de Spearman	60
4.2.2. Conjunto de datos	61
4.2.3. Tipos de bacterias asociadas con la VB	62
4.2.4. Análisis de trayectorias	62
4.2.5. Métricas estadísticas en el análisis de trayectorias	63
4.2.6. Clasificación de los efectos totales en el análisis de trayectorias	63
4.3. Estudio experimental	64
4.4. Resultados	73
4.5. Discusión	74
4.6. Conclusiones	75
5. Conclusiones y recomendaciones generales	82
5.1. Conclusiones generales	82
5.2. Aportaciones de la investigación	83
5.3. Limitaciones del estudio	84
5.4. Recomendaciones	85
5.5. Publicaciones derivadas de la investigación	85

Anexo

87

Universidad Juárez Autónoma de Tabasco.
México.

Índice de tablas

1.1. Lista de actividades.	10
2.1. Métricas estadísticas.	18
2.2. Datos imputados con la moda del atributo categórico de Anticoncepcion y diagnóstico de VB.	21
2.3. Datos imputados con media del atributo numérico InicioSexual y diagnóstico de VB.	22
3.1. Métricas estadísticas para evaluar el ajuste del modelo en el análisis de trayectorias	36
3.2. Resumen de la estructura de los datos.	40
3.3. Datos imputados con la media del atributo <i>N.EMBARAZO</i> .	42
3.4. Datos imputados con la moda del atributo <i>FlujoV</i> .	42
3.5. Varianza de variables ilustrativas en el conjunto de datos.	44
3.6. Prueba de Kolmogorov-Smirnov (<i>p</i> -values) antes y después de las transformaciones aplicadas a cada variable ilustrativa.	45
3.7. Valores de las métricas del modelo.	48
3.8. Efectos totales de las bacterias con <i>DxVBNoMh</i> (DVB). Con base en la Figura 3.9, el efecto calculado se obtiene mediante la suma algebraica del efecto directo con el efecto indirecto, el cual a su vez se obtiene multiplicando los coeficientes de las flechas.	50
4.1. Métricas estadísticas utilizadas para evaluar el ajuste del modelo en el análisis de trayectorias.	63
4.2. Clasificación de la magnitud de los efectos totales según Hair et al. Hair et al., 2019.	64

4.3. Valores de las métricas para el modelo (4.2) y el modelo (4.3).	68
4.4. Efectos totales de las bacterias sobre DxVBNoMh (DVB) en cada etapa.	70
4.5. Interpretación de los efectos totales de cada bacteria sobre el diagnóstico VB- e Indeterminado.	71
4.6. Interpretación de los efectos totales de cada bacteria en el diagnóstico Indeterminado y VB+.	72

Universidad Juárez Autónoma de Tabasco.
México.

Índice de figuras

2.1. Nomenclatura de modelo de ecuaciones estructurales.	17
2.2. Conjuntos de matrices de correlación.	23
2.3. Matriz de correlación de los atributos relevantes.	23
2.4. Modelos gráficos causales simples. Abreviaturas: BVA representa BVAB2 y DVB representa DxVBNoMh.	24
2.5. Modelos gráficos causal de las variables relevantes.	26
3.1. Representación gráfica del modelo AT, que ilustra las relaciones causales entre variables mediante efectos y asociaciones observables, adaptado de Brand, 2021; Civelek, 2018.	35
3.2. Etapas aplicadas durante el desarrollo del análisis.	39
3.3. Diagramas de caja y bigotes de <i>Glu</i> , <i>Col</i> , <i>Trig</i> y <i>Hgb</i> , mostrando la distribución y variabilidad de los niveles de glucosa, colesterol, triglicéridos y hemoglobina.	41
3.4. Metodología aplicada para el análisis de trayectorias: desde la exploración es- tadística hasta la evaluación del modelo gráfico causal.	43
3.5. Gráficos Q-Q de <i>SemanasGesta</i> , evaluando la distribución de las semanas de ges- tación en comparación con una distribución normal teórica.	45
3.6. Gráficos Q-Q de <i>Av</i> , evaluando la distribución de <i>Atopobium vaginae</i> en compara- ción con una distribución normal teórica.	46
3.7. Matriz de correlación de <i>DxVBNoMh</i> (categorías de diagnóstico de VB: VB+, Inde- terminado, VB-) y presencia bacteriana, destacando sus asociaciones.	47

3.8. Matriz de correlación de <i>DxVBNoMh</i> (categorías de diagnóstico de VB: VB+, VB-) y presencia bacteriana, destacando sus asociaciones.	48
3.9. Modelo gráfico causal que ilustra las relaciones entre atributos bacterianos predictivos y el diagnóstico de VB.	49
4.1. Metodología del estudio	64
4.2. Matriz de correlación entre <i>DxVBNoMh</i> (categorías diagnósticas: VB- e I) y la presencia bacteriana, destacando sus asociaciones.	66
4.3. Matriz de correlación entre <i>DxVBNoMh</i> (categorías diagnósticas: I y VB+) y la presencia bacteriana, destacando sus asociaciones.	67
4.4. Modelo gráfico causal que representa los vínculos entre los atributos bacterianos predictivos y el diagnóstico en las categorías VB- e I. Las líneas continuas denotan efectos directos/indirectos; las líneas discontinuas denotan covariaciones.	69
4.5. Modelo gráfico causal que representa los vínculos entre los atributos bacterianos predictivos y el diagnóstico en las categorías I y VB+. Las líneas continuas denotan efectos directos/indirectos; las líneas discontinuas denotan covariaciones.	70

Resumen

La vaginosis bacteriana (VB) es una alteración de la microbiota vaginal caracterizada por un desequilibrio en las comunidades bacterianas, cuya composición puede variar entre mujeres y evolucionar a través de distintos estados. Comprender las interacciones entre las bacterias asociadas y su influencia en el desarrollo de la VB es fundamental para mejorar su interpretación clínica y biológica. En esta investigación se analizan dichas relaciones mediante la aplicación del Análisis de Trayectorias (AT), a partir del cual se construyen Modelos Gráficos Causales (MGC), que permiten representar y estudiar las relaciones entre especies bacterianas involucradas en esta infección. El análisis se realizó con datos de 132 mujeres embarazadas entre 4 y 24 semanas de gestación. Inicialmente se construyó un modelo basado en la matriz de correlación de Pearson para identificar bacterias asociadas al diagnóstico de VB. Posteriormente, se desarrollaron MGC basados en la correlación de Spearman para examinar las relaciones causales entre bacterias y analizar las transiciones entre microbiota normal (VB-), estado intermedio (I) y VB positiva (VB+). Los resultados destacan la participación de *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera* Tipo 1 (MT1) y Bacteria Asociada con Vaginosis Bacteriana Tipo 2 (BVAB2). Además, se observó que MT1 contribuye a la transición desde VB- hacia el estado I, mientras que Av y BVAB2 influyen en la transición desde I hacia VB+. Estos modelos contribuyen a comprender mejor la dinámica bacteriana asociada a la VB.

Palabras clave: Análisis de Trayectorias, Matriz de Correlación, Métricas Estadísticas, Bacterias Anaerobias

Abstract

Bacterial vaginosis (BV) is a condition characterized by an imbalance in the vaginal microbiota, in which the composition of bacterial communities may vary among women and evolve through different states. Understanding the interactions among BV-associated bacteria and their influence on disease development is essential for improving their clinical and biological interpretation. In this research, these relationships are examined using Path Analysis (PA), from which Causal Graphical Models (CGMs) are constructed. These models enable the representation and analysis of relationships among bacterial species involved in this vaginal infection.

The analysis was conducted using data from 132 pregnant women between 4 and 24 weeks of gestation. Initially, a model based on the Pearson correlation matrix was developed to identify bacteria associated with BV diagnosis. Subsequently, CGMs based on Spearman correlations were constructed to examine causal relationships among bacteria and to analyze transitions between normal microbiota (BV⁻), an intermediate state (I), and the BV-positive state (BV⁺).

The results highlight the involvement of *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera* Type 1 (MT1), and Bacteria Associated with Bacterial Vaginosis Type 2 (BVAB2), evidencing their influence on BV diagnosis. Furthermore, MT1 was found to contribute to the transition from BV⁻ to the intermediate state, whereas Av and BVAB2 influence the progression from I to BV⁺. Overall, the proposed models contribute to a deeper understanding of the bacterial dynamics associated with BV.

Keywords: Path Analysis; Correlation Matrix; Statistical Metrics; Anaerobic Bacteria.

Capítulo 1

Protocolo de tesis

1.1. Introducción

La importancia de la investigación científica en medicina radica en su capacidad para obtener conocimientos relevantes que contribuyen al diagnóstico y tratamiento de enfermedades. En el campo de la salud, se distinguen tres tipos principales: investigación biomédica, investigación clínica e investigación en salud pública (Mendoza-Sánchez et al., 2023). La aplicación de Aprendizaje Automático (AA) en la predicción de enfermedades no es algo reciente o novedoso, dado que se ha utilizado por ejemplo para predecir si una persona tiene cáncer de mama, diabetes, tuberculosis, covid-19, entre otros. Una de las enfermedades que comúnmente afecta a las mujeres en edad fértil es la vaginosis bacteriana (VB), la cual es una afección que muy poco se ha estudiado mediante el uso de la tecnología.

La vaginosis bacteriana (VB) es una de las infecciones más frecuentes en la mujer, esto se debe a la disminución de lactobacilos en la flora vaginal. Esta infección provoca que las mujeres tengan secreciones anormales, picazón y mal olor. Las primeras investigaciones se realizaron en 1892 cuando el investigador alemán Krönig publicó dibujos de secreciones vaginales usando la prueba tinción de Gram en mujeres, considerando a los estreptococos anaerobios como causantes. En 1955, Gardner y Dukes, realizaron estudios en donde nombraron a esta enfermedad como Haemophilus Vaginalis (Romero-Herrero y Andreu-Domingo, 2016). Actualmente se conoce como Vaginosis Bacteriana y Haemophilus vaginalis la denominaron como Gardnerella vaginalis, en reconocimiento a la labor de Garner. Por lo anterior, la VB es considerada como una enfermedad de

impacto significativo en la salud de las mujeres.

Conforme han avanzado las investigaciones sobre la VB, se ha determinado que existen otras bacterias aparte de la *Gardnerella vaginalis*, como son la *Prevotella*, *Porphyromonas*, *Bacteroides*, *Peptostreptococcus*, *Mycoplasma hominis* y *Ureaplasma urealyticum*. A pesar de que la mayoría de las mujeres adquieren la VB, se ha observado que la enfermedad es asintomática, porque entre el 50 % – 75 % no presentan síntomas. La mayoría de los especialistas utilizan los criterios de Amsel (Salas et al., 2022) o la Puntuación de Nugent, para dar un diagnóstico clínico a las mujeres que presentan secreciones vaginales. El método que más usa es el de Amsel, esto se debe a que los resultados se obtienen en términos de parámetros clínicos, en cambio el método de Puntuación de Nugent no se aplica frecuentemente porque requiere del estudio tinción de Gram y se basa en parámetros microbiológicos.

Para realizar un diagnóstico clínico de la VB, se debe tener tres de los cuatro criterios de Amsel:

- Descarga vaginal homogénea fina y blanco-grisácea
- pH vaginal > 4.5
- Test de aminas (+)
- Presencia de más de 20 % de células clave con lente de alto poder en montaje al fresco de la descarga vaginal

En Sanchez-Garcia et al., 2019 los autores introducen el criterio molecular para el diagnóstico de la vaginosis bacteriana en mujeres del estado de Tabasco, quienes eran activas sexualmente y con edades comprendidas entre los 18 y 50 años. Estas mujeres acudieron para someterse a una revisión ginecológica de rutina. El estudio involucró la toma de muestras cervicovaginales, las cuales fueron analizadas mediante técnicas de PCR (Reacción en Cadena de la Polimerasa) y secuenciación de ADN (Ácido Desoxirribonucleico).

En Morelli-Martínez y Gamboa-Miranda, 2022 los autores indican que de los diferentes factores que influyen en la VB, uno corresponde a las variaciones hormonales que producen pacientes con el embarazo, las cuales son capaces de afectar la microbiota vaginal. Esta afección puede causar graves consecuencias para la salud de la madre y el feto. Normalmente, la prevalencia de

VB en mujeres embarazadas es de un 5.8 a 19.3%; sin embargo, de este porcentaje, hasta un 50 % de los casos no presenta manifestaciones clínicas.

Los avances tecnológicos (Benavent-Nuñez et al., 2020) son de gran relevancia en el área de la salud. Por ejemplo, actualmente los registros de los pacientes se tienen de forma electrónica, esto permite que el conjunto de datos se pueda analizar e interpretar, usando Aprendizaje Automático (AA) que es una rama de la inteligencia artificial (IA) para encontrar patrones significativos y así ayudar a los médicos en la toma de decisiones diagnósticas.

También Álvarez-Vega et al., 2020, consideran que las tecnologías tienen el potencial de mejorar los sistemas de salud al analizar grandes cantidades de datos clínicos para crear modelos predictivos, de tamizaje y diagnósticos. Además, destaca que el uso de estas herramientas puede mejorar el acceso a servicios médicos, reducir consultas innecesarias, brindar comodidad a los usuarios y reducir la carga de trabajo y los costos. De aquí que la inteligencia artificial y el aprendizaje automático son importantes en el campo de la medicina.

Por otro lado, un gráfico estadístico (Sagaró-del-Campo y Zamora-Matamoros, 2019) facilita la comprensión y comunicación de la evidencia proporcionada por los datos. Su utilidad es dual, ya que no solo se utiliza para presentar información que ha sido previamente organizada en una tabla, sino también para llevar a cabo un análisis exploratorio de los datos. En algunos casos, se convierte en un paso esencial durante el procesamiento de la información. Los gráficos que indican causalidad se fundamentan en la teoría de grafos, posibilitando la representación visual de una gran cantidad de nodos interconectados que reflejan las relaciones entre el conjunto de variables.

Existen diversas técnicas (Sucar, 2021) de modelado causal, por ejemplo: ecuaciones funcionales, diagramas de trayectoria o ruta, modelos de ecuaciones estructurales y redes bayesianas, para analizar y representar estas relaciones en distintos contextos. Los modelos de ecuaciones estructurales (MEE) (Iglesias Labraca y Reche Lorite, 2021) representan una herramienta estadística multivariada cuyo propósito es validar y cuantificar las relaciones causales entre diversas variables. La relevancia de los MEE radica en su capacidad para especificar vínculos entre variables que son tanto observables como no observables (latentes). Los MEE se dividen en tres tipos: modelos de variables observables (análisis de trayectoria), análisis factorial y MEE general.

El análisis de trayectorias (Lepera, 2021), examina tanto los efectos directos como los indirectos

tos de un conjunto de variables observables sobre otras, a través del estudio de sus correlaciones. Tal análisis se plasma visualmente utilizando el denominado diagrama de trayectorias, que es una representación gráfica mediante grafos, siguiendo símbolos y convenciones específicos.

En la sección del estado del arte, se presentan algunos estudios relacionados con la aplicación de los MEE, los cuales están enfocados en el área de la salud. Tales documentos serán la base para desarrollar la propuesta de investigación, en la cual se hará uso de los MEE, en particular, el análisis de trayectoria, para diseñar el modelo gráfico causal.

1.2. Marco teórico

1.2.1. Inteligencia artificial

Es la rama de la informática cuyo objetivo es la simulación de tareas tradicionalmente relacionadas con la inteligencia humana por parte de máquinas. Incluye procesos como el aprendizaje (integración de nueva información y reglas para su manejo), el razonamiento o la mejora mediante retroalimentación Benavent-Nuñez et al., 2020.

1.2.2. Aprendizaje automático

Es una tecnología que utiliza técnicas estadísticas y algoritmos computacionales para proporcionar a los ordenadores la capacidad de aprender, es decir, mejorar sus resultados en una tarea específica tras procesar datos en suficiente cantidad y sin instrucciones explícitas externas proporcionadas por el programador Álvarez-Vega et al., 2020.

1.2.2.1. Aprendizaje supervisado

Estos algoritmos buscan predecir un resultado (variable dependiente) mediante el análisis de un conjunto de características (variables independientes). Para lograr esto, emplean conjuntos de datos con etiquetas para entender la relación entre las características y los resultados, permitiéndoles realizar predicciones en conjuntos de datos nuevos y no etiquetados Harrison y Sidey-Gibbons, 2021.

1.2.2.2. Aprendizaje no supervisado

Estos algoritmos tienen como objetivo descubrir patrones previamente no definidos en conjuntos de datos, por ejemplo, agrupando observaciones similares en clústeres. Se basan en datos que no han sido previamente etiquetados por un supervisor humano, es decir, observaciones que no han sido categorizadas de antemano Harrison y Sidey-Gibbons, [2021](#).

1.2.3. Big data

Se define como el conjunto de métodos informáticos empleados para examinar una gran cantidad de datos con el objetivo de identificar información que revele patrones, conexiones y asociaciones significativas para una entidad organizativa Márquez-Díaz, [2020](#). El concepto de Big Data se distingue por tres atributos fundamentales en el contexto de la investigación, denominados "Volumen", "Variedad", y "Velocidad", reconocidos como las tres V del Big Data.

1.2.4. Lenguaje de programación R

R es un lenguaje de programación empleado como software estadístico en diversas disciplinas Jahuey-Martinez et al., [2022](#). Gracias a su gratuidad y facilidad de uso, se ha consolidado como la plataforma estadística más empleada para el análisis de datos y big data. Se encuentra a la par de otros lenguajes y programas estadísticos como Python, Julia, SPSS y SAS. Puede ser utilizado a través de distintas plataformas, tales como R, RStudio, RStudio Cloud y Microsoft R Open.

1.2.5. Vaginosis bacteriana (VB)

Es un síndrome clínico polimicrobiano, resultado de la sustitución de las especies de *Lactobacillus* productoras de peróxido de hidrógeno por altas concentraciones de bacterias anaeróbicas (p. ej., *Prevotella* sp., *Mobiluncus* sp., y *Atopobium vaginae*), *Gardnerella vaginalis*, *Mycoplasma hominis*, *Ureaplasma urealyticum* y numerosos anaerobios fastidiosos o no cultivables Romero-Herrero y Andreu-Domingo, [2016](#).

1.2.5.1. Vaginosis bacteriana en embarazadas

La infección vaginal N. González et al., 2019 es una enfermedad de distribución mundial convertida en un problema de salud debido al aumento en su incidencia, constituyendo una de las complicaciones médicas más frecuentes asociadas al embarazo y por las consecuencias que trae para la salud de la madre y su producto. Durante el embarazo, ocurren marcados cambios del medio hormonal, por lo cual se produce alteración en la flora vaginal, haciendo de las infecciones cérvico-vaginales un trastorno frecuente durante el embarazo y a menudo difícil de erradicar, produciéndose recurrencias de las mismas.

1.2.6. Modelos de ecuaciones estructurales (MEE)

Se trata de un conjunto de técnicas pertenecientes a la Estadística Multivariante que posibilitan el análisis de la relación de dependencia o causalidad entre variables observables y/o variables no observables conocidas como latentes Lepera, 2021.

1.2.6.1. Tipos de modelos de ecuaciones estructurales

Para realizar una síntesis adecuada de los modelos de ecuaciones estructurales Iglesias Labra y Reche Lorite, 2021, es necesario distinguir entre tres categorías:

1. Los modelos de variables observadas, también llamados análisis de trayectorias, excluyen variables latentes y suponen la inexistencia de errores de medición. Estos modelos son una ampliación del modelo de regresión múltiple, ya que consideran múltiples ecuaciones de regresión que pueden estimarse simultáneamente.
2. El análisis factorial es una técnica estadística de reducción de datos utilizada para explicar las correlaciones entre variables observadas utilizando el menor número posible de variables latentes. Se diferencia entre análisis factorial confirmatorio y análisis factorial exploratorio según si el objetivo es confirmar una estructura factorial previamente definida o descubrirla en caso de que no se haya considerado ninguna.
3. El modelo de ecuaciones estructurales general es una generalización de los dos modelos mencionados anteriormente. Permite evaluar las relaciones entre variables observadas y

latentes, así como las relaciones entre constructos.

1.3. Justificación

En los últimos años, la literatura científica ha mostrado una limitada producción relacionada con el estudio de la vaginosis bacteriana. La mayoría de las publicaciones disponibles se centran en describirla como una afección ginecológica, abordando principalmente su impacto en mujeres en edad fértil, los factores asociados a su aparición, así como las complicaciones que puede generar durante el embarazo tanto en la madre como en el recién nacido.

Para investigar sobre la VB, De la Cruz-Ruiz et al., 2024 y Hernández-Gómez et al., 2023 analizaron un dataset que constaba de mujeres sexualmente activas del estado de Tabasco con registros de Vaginosis Bacteriana. El dataset se clasificó en tres categorías: vaginosis positiva, vaginosis negativa y vaginosis indeterminada. La investigación se desarrolló mediante algoritmos orientados a identificar las bacterias coexistentes más prevalentes para la VB, con el algoritmo a priori se generaron reglas de asociación para diagnosticar VB positiva y con el algoritmo agglomerative hierarchical clustering se conformaron grupos por alta intrasimilitud e interdisimilitud.

Basado en investigaciones documentales y aprovechando el uso de tecnologías computacionales, el presente estudio propuso un enfoque alternativo para el diagnóstico de la vaginosis bacteriana (VB). Se construyeron modelos gráficos causales a partir de un conjunto de datos de mujeres embarazadas con diagnóstico de VB, con el objetivo de representar de manera visual la influencia de las bacterias coexistentes en este grupo poblacional del estado de Tabasco. La finalidad de este enfoque es proporcionar a los médicos una referencia clara sobre las bacterias implicadas, facilitando así la toma de decisiones adecuadas respecto al tratamiento requerido por las mujeres embarazadas y contribuyendo a la mejora de su salud.

1.4. Pregunta de investigación

¿Determina el modelo gráfico causal la influencia de cada una de las bacterias coexistentes detonadoras de vaginosis bacteriana en mujeres embarazadas?

1.5. Hipótesis

El modelo gráfico causal representa la influencia de cada una de las bacterias coexistentes detonadoras de vaginosis bacteriana en mujeres embarazadas con una precisión de al menos el 80 %.

1.6. Objetivo general

Diseñar un modelo gráfico basado en análisis de trayectorias que represente la influencia de las bacterias coexistentes detonadoras de vaginosis bacteriana en mujeres embarazadas.

1.6.1. Objetivos específicos

- Diseñar un modelo gráfico que represente las relaciones causales entre las bacterias y con el diagnóstico de vaginosis bacteriana positivo.
- Con base en el modelo gráfico causal determinar la influencia de cada una de las bacterias coexistentes detonadoras de vaginosis bacteriana en mujeres embarazadas.

1.7. Metodología

Para el desarrollo de la investigación se realizó de acuerdo a lo siguiente

- Definición del modelo a usar: Los modelos gráficos causales que se implementarán en este estudio se basó en el Análisis de Trayectorias (AT), una herramienta integral de los Modelos de Ecuaciones Estructurales. En esta fase, se delineó la estructura del modelo, identificando las trayectorias causales entre las variables de interés.
- Publicaciones especializadas de los últimos años: Investigación sobre la literatura existente del la aplicación del AT que están enfocados en el área de la salud, en diferentes fuentes de información.

- Exploración del dataset en R: Se llevó a cabo mediante el uso del lenguaje de programación R. En este periodo, se empleo diversas técnicas y funciones en R para analizar, visualizar y comprender los datos contenidos en el conjunto de datos.
- Preprocesamiento del dataset en R: Consistió en la aplicación de técnicas de limpieza y transformación de datos mediante el uso del lenguaje de programación R. Durante esta etapa, se llevaron a cabo diversas operaciones para garantizar la calidad y la preparación adecuada de los datos antes de realizar el estudio.
- Diseñar el modelo gráfico causal en R: En este proceso, se implemento las estructuras y relaciones causales definidas previamente en el AT. Se emplearon las funciones y paquetes específicos de R dedicados a la construcción y visualización de modelos gráficos causales, asegurando así una representación clara y precisa de las interacciones entre las variables del estudio.
- Realización de pruebas con el dataset: Se aplicó el modelo gráfico causal previamente diseñado, para ello, se evaluó la efectividad del modelo al utilizar el conjunto de datos, verificando la coherencia y validez de las relaciones causales propuestas.
- Análisis y comparación de los resultados: Este análisis se llevó a cabo mediante una comparación con un modelo gráfico alternativo (redes bayesianas). Además, se verifico los resultados obtenidos consultando con un biólogo experto en el tema de estudio. La combinación de estas evaluaciones permitió una comprensión más completa y robusta de los hallazgos del estudio.

1.8. Cronograma de actividades

Tabla 1.1. Lista de actividades.

No.	Actividad	2023		2024		2025	
		S1	S2	S3	S4	S5	S6
1	Revisión de literatura	✓	✓	✓	✓		
2	Redacción y presentación del protocolo	✓	✓				
3	Definición del modelo gráfico causal	✓	✓	✓			
4	Exploración del conjunto de datos		✓	✓			
5	Preprocesamiento del conjunto de datos			✓	✓		
6	Diseñar el modelo gráfico causal			✓	✓		
7	Realizar pruebas con el modelo diseñado				✓	✓	
8	Análisis y comparación de resultados					✓	✓
9	Redacción y envío de artículo				✓	✓	✓
10	Redacción y revisión de la tesis			✓	✓	✓	✓
11	Presentación y defensa de la tesis						✓

Referencias del capítulo

- Álvarez-Vega, M., Quirós-Mora, L. M., & Cortés-Badilla, M. V. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, 5(8), 1-12. <https://doi.org/10.31434/rms.v5i8.557>
- Benavent-Nuñez, D., Colomer-Mascaró, J., Quecedo-Gutiérrez, L., Gol-Montserrat, J., & Del Llano-Señaris, J. E. (2020). *Inteligencia Artificial y decisiones Clínicas*. Fundación Gaspar Casal.
- De la Cruz-Ruiz, F., Canul-Reich, J., Rivera-López, R., & De la Cruz-Hernández, E. (2024). Impact of data balancing a multiclass dataset before the creation of association rules to study bacterial vaginosis. *Intelligent Medicine*, 4(3), 188-199. <https://doi.org/10.1016/j.imed.2023.02.001>

- González, N., Santisteban, A., Ortiz, Y., Pérez, D., & González, M. (2019). Factores de riesgo asociados a infección vaginal en mujeres embarazadas. *Multimed. Revista Médica Granma*, 23(3), 430-446.
- Harrison, C. J., & Sidey-Gibbons, C. J. (2021). Machine learning in medicine: a practical introduction to natural language processing. *BMC Medical Research Methodology*, 21(1), 158. <https://doi.org/10.1186/s12874-021-01347-1>
- Hernández-Gómez, H. J., Canul-Reich, J., Hernández-Ocaña, B., & De la Cruz-Hernández, E. (2023). An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis*, 27(3), 583-611. <https://doi.org/10.3233/IDA-216488>
- Iglesias Labraca, J. M., & Reche Lorite, F. (2021). Modelos de Ecuaciones Estructurales. *Universidad de Almería*, 1-66. <https://repositorio.ual.es/bitstream/handle/10835/13177/IGLESIAS%20LABRACA,%20JUAN%20MIGUEL.pdf?sequence=1>
- Jahuey-Martinez, F. J., Herrera-Ojeda, J. B., & Paredes-Sánchez, F. A. (2022). El programa R: una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria*, 23(4). <https://doi.org/10.22201/cuaieed.16076079e.2022.23.4.4>
- Lepera, A. (2021). Introducción a los Modelos de Ecuaciones Estructurales y su Implementación en R mediante un ejemplo. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión Económica*, 1(1), 15-37. <https://www.economicas.uba.ar/wp-content/uploads/2016/04/Lepera-Andrea-1.pdf>
- Márquez-Díaz, J. (2020). Inteligencia artificial y Big Data como soluciones frente a la COVID-19. *Revista de Bioética y Derecho*, 50, 315-331. <https://doi.org/10.1344/rbd2020.50.31643>
- Mendoza-Sánchez, D., Quintana-Zúñiga, Á. A., & Díaz-Quiñonez, A. (2023). La importancia de la investigación científica en medicina. *Revista Médica Sinergia, Atención Familiar UNAM*, 24(3), 224-227. <https://doi.org/10.22201/fm.14058871p.2023.3>
- Morelli-Martínez, I., & Gamboa-Miranda, S. (2022). Vaginosis bacteriana en el embarazo: últimos avances hasta la fecha. *Revista Médica Sinergia*, 7, e838. <https://doi.org/10.31434/rms.v7i7.838>
- Romero-Herrero, D., & Andreu-Domingo, A. (2016). Vaginosis Bacteriana. *Enfermedades Infecciosas y Microbiología Clínica*, 34(3), 14-18.

- Sagaró-del-Campo, N. M., & Zamora-Matamoros, L. (2019). Métodos gráficos en la investigación biomédica de causalidad. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, 44(4), 1-9. <http://revzoilomarinaldo.sld.cu/index.php/zmv/article/view/1846>
- Salas, J., Angulo, L., & Mendez, E. (2022). Vaginosis bacteriana: actualización y novedad terapéutica. *UCIMED*, 5(6), 85-93. <https://doi.org/10.34192/cienciaysalud.v5i6.387>
- Sanchez-Garcia, E. K., Contreras-Paredes, A., Martinez-Abundis, E., Garcia-Chan, D., Lizano, M., & De la Cruz-Hernandez, E. (2019). Molecular epidemiology of bacterial vaginosis and its association with genital micro-organisms in asymptomatic women. *Journal of Medical Microbiology*, 68(9), 1373-1382. <https://doi.org/10.1099/jmm.0.001044>
- Sucar, L. E. (2021). *Probabilistic Graphical Models: Principles and Applications* (2nd). Springer.

Capítulo 2

Hacia un modelo gráfico causal para el análisis de la vaginosis bacteriana

Universidad Juárez Autónoma de Tabasco.
México.

Hacia un modelo gráfico causal para el análisis de la vaginosis bacteriana

Maricela García Avalos¹; Juana Canul Reich¹; Lil María Xibai Rodríguez-Henríquez²; Erick Natividad De la Cruz Hernández¹.

Resumen El Análisis de Trayectorias (AT) es una técnica de Estadística Multivariante que estudia los efectos directos (causales) e indirectos de un conjunto de variables observables a través del análisis de correlaciones. Este estudio aplica AT a un conjunto de datos sobre la Vaginosis Bacteriana (VB) en mujeres embarazadas, el cual tiene tres tipos de diagnósticos. Aunque se han realizado investigaciones sobre las bacterias asociadas a la VB utilizando métodos como selección de variables, agrupaciones y reglas de asociación, el AT no ha sido utilizado como herramienta visual para detectar las variables coexistentes que varían de una mujer a otra. Se presenta un Modelo Gráfico Causal (MGC) que se basa en un modelo teórico derivado de la matriz de correlación de Pearson y métricas estadísticas. Este modelo muestra las variables que influyen en el diagnóstico de VB, identificando las bacterias relevantes: Mycoplasma Hominis (Mh), Atopobium Vaginae (Av), Gardnerella Vaginalis (Gv), Megasphaera Tipo 1 (MT1) y Bacteria Asociada a Vaginosis Bacteriana Tipo 2 (BVBA2). La validez biológica del modelo fue evaluada por un experto en biología. El análisis se llevó a cabo utilizando el software R.

Palabras clave: Vaginosis Bacteriana, Mujeres Embarazadas, Métricas Estadísticas, Correlación, Análisis de Trayectoria

Institución de adscripción: ¹ Universidad Juárez Autónoma de Tabasco, ² Instituto Nacional de Astrofísica, Óptica y Electrónica.

2.1. Introducción

La investigación científica en medicina es importante para mejorar el diagnóstico y tratamiento de enfermedades. Se divide en tres tipos: investigación biomédica, que estudia los mecanismos biológicos; investigación clínica, que evalúa intervenciones en pacientes; e investigación en salud pública, que analiza factores sociales que afectan la salud (Mendoza-Sánchez et al., 2023). La VB es una afección común en mujeres en edad fértil, especialmente en aquellas sexualmente activas, y puede presentarse de forma asintomática o sintomática. El síntoma más destacado en las mujeres con síntomas es el flujo vaginal anormal, que puede variar en color (gris o verde) y presenta un olor característico a pescado (Sanchez-Garcia et al., 2019). En biología, se ha observado que múltiples bacterias son responsables de la aparición de la VB, y que la composición de estas bacterias puede variar entre mujeres. Este estudio tiene como objetivo identificar las variables coexistentes que influyen en el diagnóstico de la VB. Esta enfermedad puede ser influenciada por fluctuaciones hormonales durante el embarazo, que alteran la microbiota vaginal. La VB en este período se asocia con un mayor riesgo de complicaciones obstétricas, como parto prematuro, ruptura prematura de membranas y endometritis puerperal. Por ello, es crucial monitorear y tratar esta condición en mujeres embarazadas para reducir los riesgos obstétricos (Morelli-Martínez y Gamboa-Miranda, 2022).

La VB puede aumentar la susceptibilidad a infecciones de transmisión sexual, como el Virus del Papiloma Humano (VPH). Diversos estudios, tanto transversales como longitudinales, han investigado la relación; sin embargo, no han encontrado asociaciones significativas en estudios transversales entre la VB y VPH (Sanchez-Garcia et al., 2019).

En el estudio de la VB en mujeres sexualmente activas se han utilizado diversos enfoques metodológicos. Uno de ellos es la selección de atributos (Pérez-Gómez, Canul-Reich, Hernández-Torruco y Hernández-Ocaña, 2020), que identifica las variables más relevantes para predecir un diagnóstico positivo. Otro enfoque es el análisis de agrupaciones (Hernández-Gómez et al., 2023), que determina grupos de bacterias asociadas con el diagnóstico, observando variaciones entre estos grupos. Además, se han aplicado reglas de asociación para identificar factores coexistentes que contribuyen al estado positivo de la VB (De la Cruz-Ruiz et al., 2024). Estos métodos buscan aclarar las variables bacterianas que influyen en el diagnóstico de esta condición.

Los autores Álvarez-Vega et al., 2020 consideran que las tecnologías tienen el potencial de mejorar los sistemas de salud al analizar grandes cantidades de datos clínicos para crear modelos predictivos, de tamizaje y diagnósticos. Los gráficos estadísticos (Sagaró-del-Campo y Zamora-Matamoros, 2019) son herramientas indispensables en el análisis exploratorio de datos en las ciencias biomédicas, ya que ayudan a identificar relaciones entre múltiples variables. Por otro lado, los Modelos de Ecuaciones Estructurales (MEE) en estadística multivariante nos permiten comprender las relaciones entre variables observables y latentes (variables no observables). El Análisis de Trayectorias es una técnica derivada del MEE que explora los efectos directos (causales) e indirectos de las variables observables y permite comprender a través de un MGC, las interacciones entre múltiples variables del sistema (Lepera, 2021).

En este trabajo se utilizó un conjunto de datos de mujeres embarazadas de zonas rurales y urbanas que acudieron a una revisión ginecológica, donde se realizaron la prueba para el diagnóstico de VB. Se aplica el AT a estos datos para obtener un MGC que ayude a visualizar las variables coexistentes que influyen en el diagnóstico de VB.

2.2. Materiales y métodos

2.2.1. Matriz de correlación

Una matriz de correlación es una representación tabular que contiene los coeficientes de correlación entre distintas variables que indica la fuerza y dirección de la relación entre pares de variables. La correlación oscila entre -1 y 1, (Molina-Arias et al., 2021) donde:

- Un valor de 1 indica una correlación positiva perfecta, lo que significa que las variables tienden a aumentar juntas.
- Un valor de -1 indica una correlación negativa perfecta lo que significa que una variable tiende a aumentar, mientras que la otra tiende a disminuir.
- Un valor de 0 indica que no existe correlación entre las variables.

2.2.2. Modelo de ecuaciones estructurales

Los MEE engloba un conjunto de técnicas estadísticas multivariantes diseñadas para examinar la relación de dependencia o causalidad entre las variables observables y latentes (no observables). La Figura 4.2, presenta la nomenclatura utilizada para representar gráficamente un MEE (Iglesias Labraca y Reche Lorite, 2021).



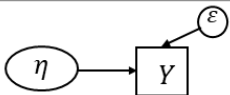

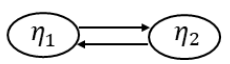
	Rectángulo o Cuadrado: significa que la variable X es observable
	Círculo o Elipse: significa que la variable η es latente
	La variable Y es un indicador de la variable η que tiene un término de error o perturbación ε
	La curva con doble flecha significa asociación entre las dos variables (covariación) ξ_1 y ξ_2
	La presencia de dos flechas simples significa que las variables tienen una causalación recíproca

Figura 2.1. Nomenclatura de modelo de ecuaciones estructurales.

Existen tres categorías de MEE: Análisis de trayectoria, análisis factorial y modelo general de ecuaciones estructurales (Barbeau et al., 2019; Iglesias Labraca y Reche Lorite, 2021). El AT es una técnica de la estadística multivariante que se centra en examinar los efectos causales e indirectos de un grupo de variables observables sobre otras variables mediante el análisis de sus correlaciones (Lepera, 2021). Dicho análisis se plasma visualmente mediante el MGC que es una representación gráfica, siguiendo símbolos y convenciones matemáticas.

2.2.3. Métricas estadísticas de MEE

En estadística, los índices de ajuste (métricas) se utilizan para evaluar el rendimiento de los modelos en conjuntos de datos observados. Los índices proporcionan una medida cuantitativa de la calidad del modelo y ayudan a determinar si el modelo describe o predice adecuadamente los datos.

La Tabla 3.1 muestra algunas de las métricas comunes utilizadas en MEE, en particular para

AT (Jordán-Muiños, 2021; Lepera, 2021), lo que nos permite verificar si el modelo teórico propuesto representa adecuadamente la estructura subyacente en los datos.

Tabla 2.1. Métricas estadísticas.

Abreviaturas	Significado	Criterio
RMSEA	Error cuadrático medio de aproximación	<0.05
SRMR	Raíz cuadrática media normalizada residual	<0.05
GFI	Índice de bondad de ajuste	≥0.95
NFI	Índice de ajuste normalizado	≥0.95
CFI	Índice de ajuste comparativo	≥0.95

Los índices RMSEA y SRMR se conocen como índices absolutos de ajuste y revelan hasta qué punto el modelo representado en la matriz de covarianzas coincide con el modelo implícito. Cuanto menor sea el resultado, mejor será la calidad del ajuste del modelo. Por otro lado, CFI, NFI y GFI, se denominan índices de ajuste incrementales y evalúan cómo mejora un modelo propuesto en comparación con un modelo base (Jordán-Muiños, 2021).

2.3. Estudio experimental

En este estudio se utilizó el lenguaje de programación R, que es un software gratuito de análisis estadístico (Jahuey-Martinez et al., 2022), que se ha consolidado como la plataforma estadística utilizada para el análisis de datos. Para el desarrollo del estudio, se inició con la exploración de la estructura de los datos, lo que implicó la eliminación de variables no relacionadas con la VB. Se abordaron también los valores atípicos, que se encontraban fuera del rango normal de cada variable. A continuación, se realizó la imputación de valores, reemplazando los datos faltantes mediante métodos estadísticos. Posteriormente, se calculó la matriz de correlación, obteniendo los coeficientes de correlación entre las variables. Se diseñó un MGC que ilustra las relaciones entre las variables, basado en la matriz de correlación, el modelo teórico y las métricas estadísticas. Finalmente, el MGC fue validado por personal del laboratorio con experiencia clínica en el área, asegurando su precisión y relevancia.

2.3.1. Conjunto de datos VB

El conjunto de datos se construyó con información de mujeres embarazadas de comunidades rurales y urbanas del estado de Tabasco, sureste de México. Las pacientes se sometieron a un estudio de diagnóstico de VB. Este conjunto de datos incluye información sociodemográfica, bacterias, el diagnóstico de VB, entre otros. La recopilación de los datos se realizó durante las campañas de embarazo saludable (agosto 2018-enero 2020), llevadas a cabo por un investigador biólogo de la División Académica Multidisciplinaria de Comalcalco (DAMC) de la Universidad Juárez Autónoma de Tabasco (UJAT).

2.3.2. Estructura del conjunto de datos

El análisis exploratorio de datos es un enfoque que utiliza estadísticas descriptivas y herramientas gráficas para obtener una comprensión profunda de un conjunto de datos. Su objetivo es identificar las características del conjunto, como la presencia de valores perdidos y atípicos, los cuales deben ser abordados para mejorar la calidad y la validez de los resultados obtenidos en el análisis posterior (Bouza, 2023). En el análisis exploratorio de datos se identificaron 87 atributos y 132 instancias, de los cuales 41 eran atributos categóricos y 46 atributos numéricos. El atributo clase (Diagnóstico de Vaginosis Bacteriana sin Mycoplasma Hominis, DxVBNoMh) se dividió en tres categorías: Vaginosis Bacteriana Positivo (VB+), que indica la presencia de la condición; Indeterminado, para casos sin un diagnóstico claro; y Microbiota Normal (VB-), que señala la ausencia de la condición. Se registraron 609 valores faltantes y 143 valores atípicos en el conjunto de datos.

2.3.3. Eliminación de atributos

Durante el análisis se identificaron variables que no están relacionadas con la VB, esto se verificó con el área clínico-biológica quien fue creador de los datos. Al tener este conocimiento se decidió eliminar tales variables, en particular se descartó todas las variables relacionadas con el VPH. Así, de los 87 atributos originales, se obtuvo 72 variables con 132 observaciones.

2.3.4. Valores atípicos

El diagrama de caja fue desarrollado por Tukey en 1977 como una herramienta para el análisis exploratorio de datos. Se utiliza para resumir y comparar distribuciones basándose en sus valores extremos (máximo y mínimo), así como en la mediana y los cuartiles. Este gráfico también se conoce como diagrama de caja y bigotes (Madrid et al., 2022). Se implementó la técnica estadística en el conjunto de datos para identificar los valores atípicos, resultando en la detección de 143 valores atípicos. A pesar de que estos datos se encuentran fuera del rango normal en las variables, corresponden a pacientes que presentan otras enfermedades. Tras consulta con un biólogo experto en el área, se determinó que era pertinente mantener estos valores originales en el conjunto de datos.

2.3.5. Valores faltantes

La imputación de datos faltantes es el proceso que utiliza la información disponible en la muestra para asignar un valor a las variables con datos ausentes. Su objetivo es crear un conjunto de datos completo para el análisis que se desea realizar. La imputación por media, que consiste en utilizar el promedio de una variable, es una técnica estadística aplicada para manejar valores faltantes en variables numéricas (Contreras-Cueva et al., 2024). Por otro lado, la imputación por moda, que se refiere a la frecuencia más común de una variable, es una técnica estadística empleada para abordar valores ausentes en variables categóricas (Mejía-Peñañiel et al., 2024).

En el análisis se detectaron 609 datos perdidos distribuidos en 47 variables, que correspondían a atributos tanto categóricos como numéricos. Para imputar los datos faltantes en las variables categóricas utilizando la moda, primero se identificó el dato ausente. Luego, se emplearon los datos restantes de esa variable, limitándose a aquellos que correspondían al diagnóstico (VB+, Indeterminado o VB-) relacionado con el dato ausente. Se calculó la moda y, finalmente, se sustituyó el valor en el dato perdido. Por ejemplo, en la Tabla 3.2 se muestran los datos imputados de las primeras diez instancias de la variable uso de anticonceptivo (Anticoncepcion). La columna DxVBNoMh indica el diagnóstico de VB, donde 1 representa VB+, 2 indica indeterminado y 3 corresponde a VB-.

Tabla 2.2. Datos imputados con la moda del atributo categórico de Anticoncepcion y diagnóstico de VB.

Instancias	Datos originales	Datos imputados	DxVBNoMh
1	NA	2	3
2	NA	2	1
3	NA	2	3
4	NA	2	3
5	2	2	3
6	NA	2	3
7	2	2	3
8	2	2	3
9	2	2	3
10	1	1	3

Para imputar los datos perdidos en las variables numéricas con la media, se siguió un proceso similar. Primero, se identificó el dato ausente y se utilizaron los datos restantes de la variable, restringiéndose a aquellos que correspondían al diagnóstico relacionado (1 representando VB+, 2 indicando indeterminado y 3 correspondiente a VB-). Luego, se calculó la media y se sustituyó el valor en el dato faltante. La Tabla 3.3 presenta los datos imputados de las primeras diez instancias del atributo inicio sexual (InicioSexual).

Tabla 2.3. Datos imputados con media del atributo numérico InicioSexual y diagnóstico de VB.

Instancias	Datos originales	Datos imputados	DxVBNoMh
1	NA	18.0137	3
2	NA	17.5000	1
3	NA	18.0137	3
4	NA	18.0137	3
5	NA	18.0137	3
6	NA	18.0137	3
7	NA	18.0137	3
8	NA	18.0137	3
9	NA	18.0137	3
10	NA	18.0137	3

2.3.6. Matriz de correlación de Pearson

Para determinar las relaciones entre las variables, se formaron varios grupos a partir del conjunto de datos, incluyendo cada uno el atributo de clase. Se calculó la matriz de correlación de Pearson para cada grupo con el objetivo de visualizar la relación de cada variable con el DxVB- NoMh (razón principal para formar los grupos). La Figura 4.3 presenta las matrices de correlación de dos de los grupos formados. En la matriz de la izquierda se resaltan cuatro variables (Av, Gv, MT1 y BVAB2) con coeficientes de correlación significativos, mientras que en la matriz de la derecha solo se observa una variable (Mh) relevante.

Después de calcular las matrices de correlación para cada grupo, se seleccionaron las variables que tenían una correlación negativa ≤ -0.39 y correlación positiva ≥ 0.54 con respecto al atributo de clase. Se formó un grupo que incluye todas las variables que cumplen con estos criterios de coeficiente de correlación. La Figura 4.4 muestra la matriz de correlación triangular superior de las variables seleccionadas (Mh, Av, Gv, MT1, BVAB2) y el DxVBNoMh. La diagonal principal de la matriz, que va de la esquina superior izquierda a la esquina inferior derecha, está formada por unos, esto indica una correlación perfecta de una variable consigo misma.

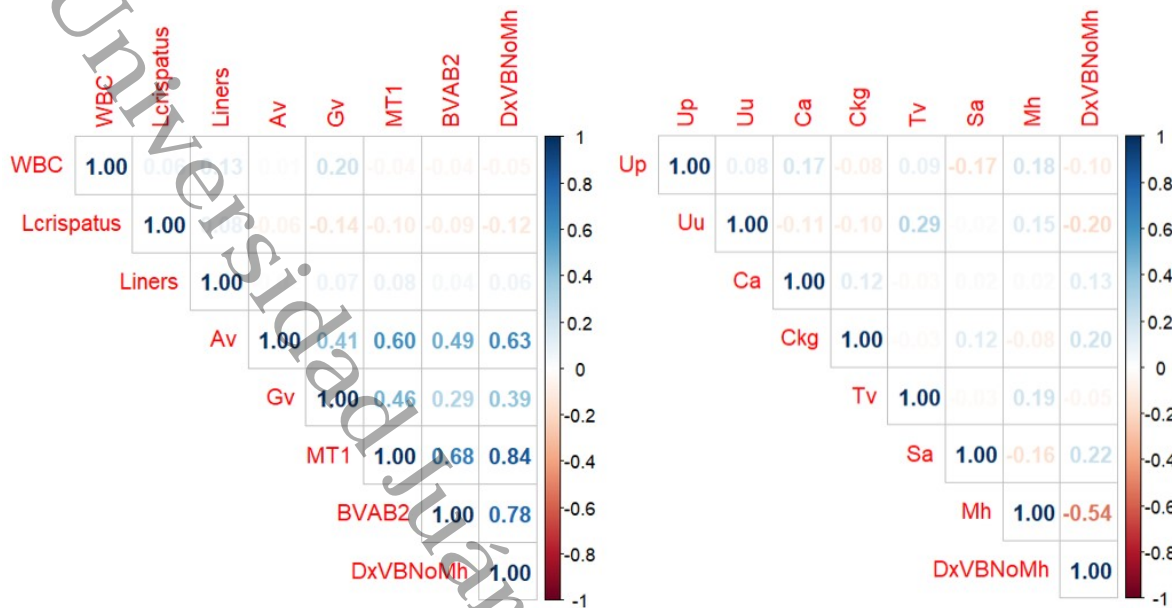


Figura 2.2. Conjuntos de matrices de correlación.

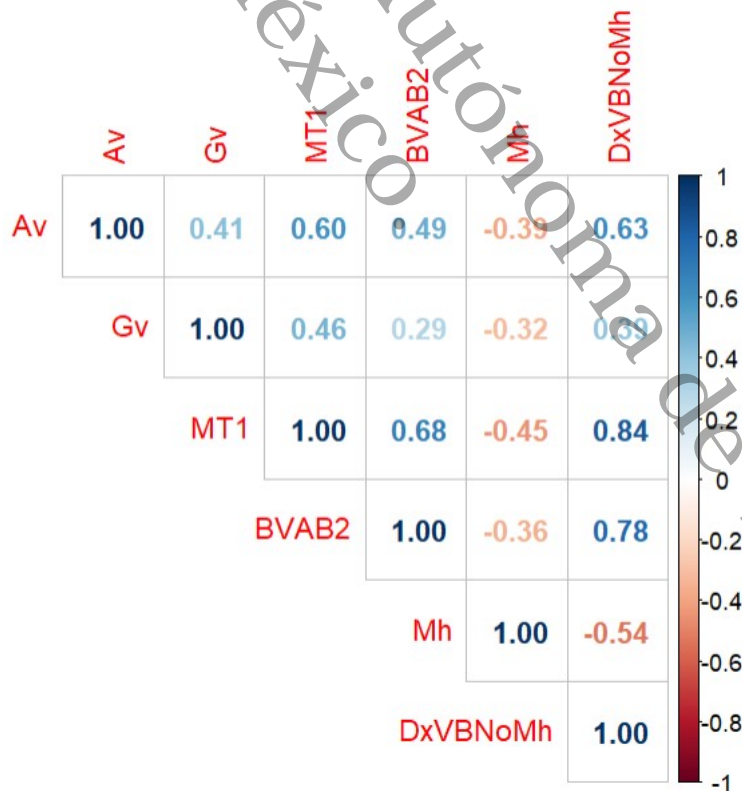


Figura 2.3. Matriz de correlación de los atributos relevantes.

2.3.7. Modelo gráfico causal

El desarrollo de un modelo gráfico causal requiere la determinación del modelo teórico basado en la matriz de correlación. Se inicia presentando modelos teóricos simples, que involucran únicamente una variable y el atributo de clase. Los modelos teóricos simples Ec. (2.1), Ec. (2.2), Ec. (2.3), Ec. (2.4) y Ec. (2.5) se representan en los diagramas a), b), c), d) y e), de la Figura 4.5, respectivamente.

$$\text{Mod1} \leftarrow' \text{DxVBNomh} \sim \text{Av}' \quad (2.1)$$

$$\text{Mod2} \leftarrow' \text{DxVBNomh} \sim \text{Gv}' \quad (2.2)$$

$$\text{Mod3} \leftarrow' \text{DxVBNomh} \sim \text{MT1}' \quad (2.3)$$

$$\text{Mod4} \leftarrow' \text{DxVBNomh} \sim \text{BVAB2}' \quad (2.4)$$

$$\text{Mod5} \leftarrow' \text{DxVBNomh} \sim \text{Mh}' \quad (2.5)$$

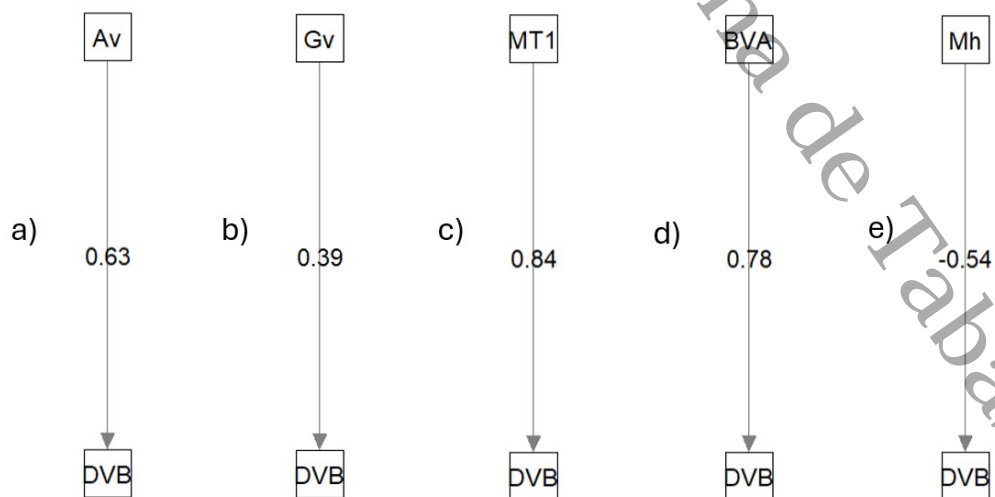


Figura 2.4. Modelos gráficos causales simples. Abreviaturas: BVA representa BVAB2 y DVB representa DxVBNomh.

Para evaluar el ajuste de los MGC simples Ec. (2.1), Ec. (2.2), Ec. (2.3), Ec. (2.4) y Ec. (2.5), se calcularon las métricas estadísticas presentadas en la Tabla 1. Los valores obtenidos de las métricas para los cinco modelos teóricos son: RMSEA = 0.000, SRMR = 0.000, GFI = 1.000, NFI = 1.000 y CFI = 1.000, todos dentro de los rangos esperados de acuerdo con la Tabla 3.1.

La Figura 4.5 muestra las relaciones causales individuales entre las variables Av, Gv, MT1, BVA y Mh respecto al diagnóstico de DVB, así como los coeficientes de las trayectorias: 0.63, 0.39, 0.84, 0.78 y -0.54, respectivamente. Los coeficientes positivos indican que, a medida que aumentan las bacterias, también incrementa la probabilidad de obtener el DVB. En cambio, el coeficiente negativo indica que la reducción de la bacteria provoca un aumento en la probabilidad de tener el diagnóstico de VB.

Después de analizar los MGC simples, se aplica el mismo AT para desarrollar un MGC que incluye las variables relevantes junto con el atributo de clase, con el fin de visualizar las relaciones que se determinan al basarnos en la matriz de correlación de la Figura 4.4. La obtención del modelo teórico Ec. (2.6), se construyó de la siguiente forma: comenzamos considerando el DxVBNomh en relación con las variables predictoras Mh, MT1, BVAB2, Av y Gv, es decir la primera relación. Luego, se crearon las demás relaciones basándose en la variable MT1 (por estar más correlacionada con el diagnóstico de VB) y las otras variables restantes, respecto al atributo clase.

$$\begin{aligned}
 \text{ModC} &\leftarrow \\
 \text{'DxVBNomh} &\sim \text{Mh} + \text{Av} + \text{Gv} + \text{MT1} + \text{BVAB2} \\
 \text{MT1} &\sim \text{Mh} \\
 \text{MT1} &\sim \text{Av} \\
 \text{MT1} &\sim \text{Gv} \\
 \text{MT1} &\sim \text{BVAB2}'
 \end{aligned}
 \tag{2.6}$$

A medida que se añadían relaciones, se calculaban las métricas estadísticas para verificar que se cumplían las condiciones enumeradas en Tabla 3.1. Los valores obtenidos de las métricas del modelo EC. (2.6), que abarca todos los modelos teóricos, son: RMSEA=0.000, SRMR=0.000, GFI=1.000, NFI=1.000 y CFI=1.000, estos valores indican que el modelo teórico está dentro de

los rangos esperados de la Tabla 3.1.

El modelo Ec. (2.6) permitió desarrollar el MGC mostrado en la Figura 2.5, que ilustra las relaciones causales o directas (líneas continuas) entre todas las variables y el diagnóstico de VB, así como las covariaciones (líneas discontinuas). Las relaciones causales desde las variables Mh, Av, Gv, BVA y MT1 hacia DVB presentan coeficientes de -0.16, 0.14, -0.04, 0.35 y 0.46, respectivamente. Además, se identifican relaciones indirectas, como la conexión entre BVA y DVB a través de MT1, donde MT1 actúa como variable intermedia, con coeficientes de 0.46. Por otro lado, aunque la variable Gv muestra una asociación débil (coeficiente es -0.04) con DVB, su reducción está relacionada con un aumento en la presencia de otras bacterias vinculadas al diagnóstico de VB.

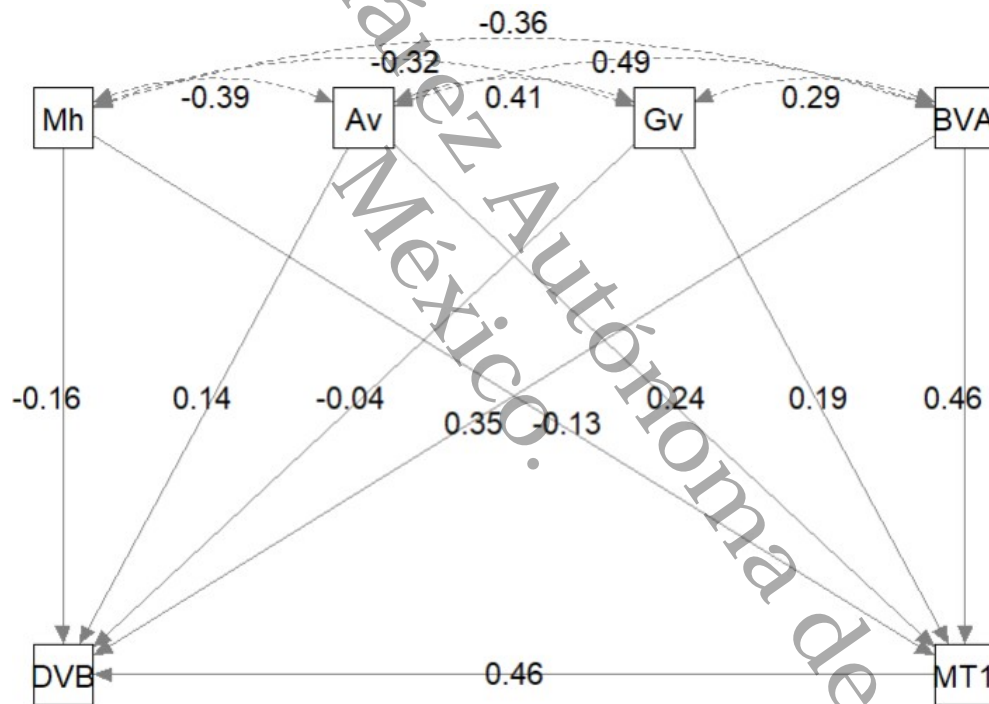


Figura 2.5. Modelos gráficos causal de las variables relevantes.

2.3.8. Validación biológica del MGC

El MGC es una herramienta visual que ilustra las relaciones entre las variables y el diagnóstico de VB, facilitando la identificación de las bacterias coexistentes Mh, Av, Gv, BVA y MT1 que influyen en el DVB, las cuales son frecuentes según la literatura Salas et al., 2022. El MGC mues-

tra que la bacteria Gv es la que presenta la menor asociación causal con el DVB, dado que su coeficiente es el más pequeño y negativo. No obstante, Gv tiene una mejor relación con otras bacterias asociadas al DVB.

2.4. Resultados

En este estudio se aplicó el método de AT debido a la presencia de variables observables en el conjunto de datos de mujeres embarazadas con diagnóstico de VB, que incluye 87 variables y 132 observaciones. El atributo de clase abarca tres categorías: Vaginosis Bacteriana, Indeterminado y Microbiota Normal. Se trabajó con 72 variables y 132 instancias. La matriz de correlación del conjunto permitió identificar relaciones significativas entre las variables y el atributo de clase, lo que facilitó la determinación de los modelos Ec. (2.1) a Ec. (2.6). Estos modelos fueron evaluados mediante métricas estadísticas, contribuyendo a la construcción de modelos gráficos causales ilustrados en las Figuras 4.5 y 2.5. Las variables obtenidas en el análisis de correlación son MT1, BVA, Av, Gv y Mh, las cuales se visualizan en los MGC con respecto a la condición. Además, se muestran las relaciones causales e indirectas entre las variables y el diagnóstico, así como los coeficientes de trayectorias.

2.5. Conclusiones

En este estudio, se utilizó el método de AT debido a que el conjunto de datos de mujeres embarazadas con diagnóstico de VB consta de variables observables. Se trabajó con 72 variables y 132 instancias. El AT se aplicó a partir del análisis de correlación, lo que permitió identificar relaciones significativas con el atributo de clase y facilitar la formulación de modelos evaluados mediante métricas estadísticas. Las variables obtenidas en el análisis de correlación MT1, BVA, Av, Gv y Mh se visualizaron en los MGC en relación con el DVB. Además, se mostraron las relaciones causales e indirectas entre las variables y el diagnóstico de VB, así como los coeficientes de trayectoria. Estas variables están asociadas a la VB, conforme a lo indicado en la literatura.

2.6. Declaración de contribución de autores y colaboradores

Maricela García: Ejecución de experimentos y redacción. **Juana Canul:** Diseño experimental y supervisión. **Lil María Rodríguez:** Diseño experimental y revisión. **Erick De la Cruz:** Validación biológica de los hallazgos.

Referencias del capítulo

- Álvarez-Vega, M., Quirós-Mora, L. M., & Cortés-Badilla, M. V. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, 5(8), 1-12. <https://doi.org/10.31434/rms.v5i8.557>
- Barbeau, K., Boileau, K., Sarr, F., & Smith, K. (2019). Path analysis in Mplus: A tutorial using a conceptual model of psychological and behavioral antecedents of bulimic symptoms in young adults. *The Quantitative Methods for Psychology*, 15(1), 38-53. <https://doi.org/10.20982/tqmp.15.1.p038>
- Bouza, C. N. (2023). Análisis exploratorio de datos univariados para la ciencia de datos. *Red Iberoamericana de Estudios Cuantitativos Aplicados*. <https://doi.org/10.13140/RG.2.2.26038.06722>
- Contreras-Cueva, A. B., Luna-Plascencia, N. R., Macías-Álvarez, P., & González-Robles, N. M. (2024). Imputación de datos faltantes en encuestas en hogares. *Memorias de la Décima Cuarta Conferencia Iberoamericana de Complejidad, Informática y Cibernética*, 310-316. <https://doi.org/10.54808/CICIC2024.01.310>
- De la Cruz-Ruiz, F., Canul-Reich, J., Rivera-López, R., & De la Cruz-Hernández, E. (2024). Impact of data balancing a multiclass dataset before the creation of association rules to study bacterial vaginosis. *Intelligent Medicine*, 4(3), 188-199. <https://doi.org/10.1016/j.imed.2023.02.001>
- Hernández-Gómez, H. J., Canul-Reich, J., Hernández-Ocaña, B., & De la Cruz-Hernández, E. (2023). An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis*, 27(3), 583-611. <https://doi.org/10.3233/IDA-216488>

- Iglesias Labraca, J. M., & Reche Lorite, F. (2021). Modelos de Ecuaciones Estructurales. *Universidad de Almería*, 1-66. <https://repositorio.ual.es/bitstream/handle/10835/13177/IGLESIAS%20LABRACA,%20JUAN%20MIGUEL.pdf?sequence=1>
- Jahuey-Martínez, F. J., Herrera-Ojeda, J. B., & Paredes-Sánchez, F. A. (2022). El programa R: una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria*, 23(4). <https://doi.org/10.22201/cuaieed.16076079e.2022.23.4.4>
- Jordán-Muñoz, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Revista de Investigación en Psicología Social*, 7(1), 66-71. <https://www.redalyc.org/articulo.oa?id=672371335005>
- Lepera, A. (2021). Introducción a los Modelos de Ecuaciones Estructurales y su Implementación en R mediante un ejemplo. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión Económica*, 1(1), 15-37. <https://www.economicas.uba.ar/wp-content/uploads/2016/04/Lepera-Andrea-1.pdf>
- Madrid, A. E., Valenzuela-Ruiz, S. M., Batanero, C., & Garzón-Guerrero, J. A. (2022). Interpretación del diagrama de caja por estudiantes universitarios de ciencias de la actividad física y deporte. *Educación Matemática*, 34(3). <https://doi.org/10.24844/em3403.10>
- Mejía-Peñañiel, E. F., Aguilar-Reyes, J. E., Bolaños-Logroño, P. F., & López-Ortega, J. R. (2024). Métodos de imputación por regresión, imputación por moda, imputación múltiple y árboles de decisión para variables categóricas en perspectiva comparada. *Revista De La Universidad Del Zulia*, 15(43), 541-555. <https://doi.org/10.46925/rdluz.43.29>
- Mendoza-Sánchez, D., Quintana-Zúñiga, Á. A., & Díaz-Quíñonez, A. (2023). La importancia de la investigación científica en medicina. *Revista Médica Sinergia, Atención Familiar UNAM*, 24(3), 224-227. <https://doi.org/10.22201/fm.14058871p.2023.3>
- Molina-Arias, M., Ochoa-Sangrador, C., & Ortega-Páez, E. (2021). Correlación, Modelos de regresión. *Evidencia Pediátrica*, 17(2), 17-25. <https://evidenciasenpediatria.es/articulo/7827/correlacion-modelos-de-regresion>
- Morelli-Martínez, I., & Gamboa-Miranda, S. (2022). Vaginosis bacteriana en el embarazo: últimos avances hasta la fecha. *Revista Médica Sinergia*, 7, e838. <https://doi.org/10.31434/rms.v7i7.838>

- Pérez-Gómez, J. F., Canul-Reich, J., Hernández-Torruco, J., & Hernández-Ocaña, B. (2020). Predictor selection for bacterial vaginosis diagnosis using decision tree and relief algorithms. *Applied Sciences*, 10(9), 3291. <https://doi.org/10.3390/app10093291>
- Sagaró-del-Campo, N. M., & Zamora-Matamoros, L. (2019). Métodos gráficos en la investigación biomédica de causalidad. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, 44(4), 1-9. <http://revzoilomarinellosld.cu/index.php/zmv/article/view/1846>
- Salas, J., Angulo, L., & Mendez, E. (2022). Vaginosis bacteriana: actualización y novedad terapéutica. *UCIMED*, 5(6), 85-93. <https://doi.org/10.34192/cienciaysalud.v5i6.387>
- Sanchez-Garcia, E. K., Contreras-Paredes, A., Martinez-Abundis, E., Garcia-Chan, D., Lizano, M., & De la Cruz-Hernandez, E. (2019). Molecular epidemiology of bacterial vaginosis and its association with genital micro-organisms in asymptomatic women. *Journal of Medical Microbiology*, 68(9), 1373-1382. <https://doi.org/10.1099/jmm.0.001044>

Capítulo 3

Modelo gráfico causal de la vaginosis bacteriana en mujeres embarazadas

Universidad Juárez Autónoma de Tabasco.
México.

Modelo gráfico causal de la vaginosis bacteriana en mujeres embarazadas

Maricela García Avalos¹; Juana Canul Reich¹; Lil María Xibai Rodríguez-Henríquez^{2,3}; Erick Natividad De la Cruz Hernández¹.

Resumen Este estudio desarrolló un Modelo Gráfico Causal (MGC) para analizar la Vaginosis Bacteriana (VB), una condición causada por un desequilibrio en la microbiota vaginal, cuya composición bacteriana varía entre mujeres. Mientras que estudios previos utilizaron selección de variables, agrupamiento y reglas de asociación para identificar bacterias asociadas a la VB, estos enfoques carecen de herramientas visuales para explorar relaciones causales y determinar cuáles son las más relevantes. En contraste, el MGC generado en este estudio permite la visualización de las bacterias asociadas y sus vínculos causales, identificando así las más influyentes. Se empleó el Análisis de Trayectorias (AT), un método estadístico de modelado de ecuaciones estructurales, para construir el MGC, con énfasis en variables observables y en la evaluación de efectos directos e indirectos mediante correlaciones y covarianzas. El AT se aplicó a datos de diagnóstico de VB en mujeres embarazadas que asistieron a consultas ginecológicas en Tabasco, México. El MGC, construido a partir de un modelo teórico basado en la matriz de correlación de Spearman, fue validado mediante métricas estadísticas y por un experto clínico-biológico. El modelo resultante resalta las bacterias que influyen en el diagnóstico de VB, específicamente *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera* Phylotype 1 (MT1) y Bacteria Asociada con Vaginosis Bacteriana Tipo 2 (BVAB2). Entre ellas, MT1 y BVAB2 mostraron la asociación más fuerte con la VB.

Palabras clave: Análisis de Trayectorias; Matriz de Correlación; Variabilidad; Métricas Estadísticas

Institución de adscripción: ¹ Universidad Juárez Autónoma de Tabasco, ² Instituto Nacional de Astrofísica, Óptica y Electrónica, ³ Secretaría de Ciencia, Humanidades, Tecnología e Innovación.

3.1. Introducción

La Vaginosis Bacteriana (VB) es un trastorno microbiológico común en mujeres en edad reproductiva, que afecta a más del 30 % de la población, con síntomas presentes en aproximadamente la mitad de los casos (Cameron et al., 2023; Salas et al., 2022). Cuando es sintomática, la VB se manifiesta como una secreción vaginal homogénea y maloliente, con un característico olor a pescado que se intensifica después de las relaciones sexuales (Morelli-Martínez y Gamboa-Miranda, 2022; Sanchez-Garcia et al., 2019).

Además del impacto en la calidad de vida de las mujeres, la VB está asociada con un desequilibrio en la microbiota vaginal, lo que puede derivar en complicaciones ginecológicas y obstétricas (Morelli-Martínez y Gamboa-Miranda, 2022). La VB no es causada por un único patógeno, sino por un consorcio de bacterias anaerobias, siendo *Gardnerella vaginalis* la especie más frecuentemente asociada (K. González et al., 2024). Este estudio tiene como objetivo identificar especies bacterianas coexistentes que influyen en el diagnóstico de VB en mujeres embarazadas.

Las técnicas de aprendizaje automático, incluyendo selección de atributos, agrupamiento y reglas de asociación, han sido utilizadas para identificar bacterias asociadas con la VB (De la Cruz-Ruiz et al., 2024; Hernández-Gómez et al., 2023; Pérez-Gómez, Canul-Reich y De la Cruz-Hernández, 2020). Estos enfoques ayudan a analizar grandes conjuntos de datos, mejorando la eficiencia diagnóstica y las estrategias de tratamiento (Álvarez-Vega et al., 2020). Sin embargo, aunque estos métodos identifican asociaciones relevantes, carecen de herramientas visuales para representar relaciones causales y determinar qué bacterias son más influyentes. Incorporar tales representaciones apoyaría a los profesionales de la salud en la realización de diagnósticos más precisos y en la definición de tratamientos adecuados para mujeres embarazadas con VB.

El Análisis de Trayectorias (AT) es un método estadístico multivariante dentro del modelado de ecuaciones estructurales que examina efectos directos e indirectos entre variables observables, utilizando correlaciones y covarianzas (Brand, 2021; Sagaró-del-Campo y Zamora-Matamoras, 2019; Sucar, 2021). En este estudio, el AT se aplicó mediante el software R (Jahuey-Martínez et al., 2022) a datos de diagnóstico de VB en mujeres embarazadas de Tabasco, México. El Modelo Gráfico Causal (MGC) resultante identificó cinco bacterias que influyen en el diagnóstico de VB: *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera*

Phylotype 1 (MT1) y Bacterias Asociadas con Vaginosis Bacteriana Tipo 2 (BVAB2). Las primeras cuatro bacterias coinciden con hallazgos previos de modelos de aprendizaje automático (De la Cruz-Ruiz et al., 2024; Hernández-Gómez et al., 2023; Pérez-Gómez, Canul-Reich y De la Cruz-Hernández, 2020).

A diferencia de técnicas anteriores, nuestro MGC proporciona un análisis causal más profundo, revelando que MT1 y BVAB2 son las bacterias más influyentes en el diagnóstico de VB, con valores de efecto total de 0.4900 y 0.5750, respectivamente. En contraste, los valores de Mh (-0.1937), Av (0.2627) y Gv (0.0482) reflejan contribuciones diferenciadas: Mh presenta una asociación inversa, Av una influencia moderada y Gv una participación baja.

Este estudio se describe en el resto del artículo de la siguiente manera: La Sección 4.2 presenta los materiales y métodos utilizados. La Sección 4.3 describe el estudio experimental, el preprocesamiento de datos y la aplicación del Análisis de Trayectorias para construir el Modelo Gráfico Causal. La Sección 4.4 reporta los resultados con énfasis en las bacterias más influyentes en el diagnóstico de la Vaginosis Bacteriana. La Sección 4.5 discute los hallazgos en relación con estudios previos. Finalmente, la Sección 4.6 expone las conclusiones del estudio.

3.2. Materiales y métodos

3.2.1. Matriz de correlación de Spearman

La correlación determina la relación entre dos variables cuantitativas, evaluando tanto la dirección como la intensidad. La covarianza mide cómo dos variables aleatorias varían conjuntamente y, mediante la estandarización, se obtienen los coeficientes de correlación (Molina-Arias et al., 2021).

Estos coeficientes oscilan entre -1 y 1, donde 0 indica ausencia de correlación. Un valor positivo representa una relación directa, mientras que un valor negativo denota una correlación inversa (Molina-Arias et al., 2021).

El coeficiente de correlación de Spearman (Apaza et al., 2022; Mendivelso y Rodríguez, 2021), a diferencia del de Pearson, no requiere normalidad ni una relación estrictamente lineal. Se calcula utilizando la Ecuación (3.1).




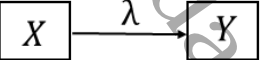
	Las variables observables se representan mediante cuadrados o rectángulos.
	Las flechas unidireccionales indican un efecto causal directo de una variable sobre otra.
	Las curvas con doble flecha significa asociación entre las dos variables (covariación).
	Los parámetros del modelo se representan sobre la flecha correspondiente.

Figura 3.1. Representación gráfica del modelo AT, que ilustra las relaciones causales entre variables mediante efectos y asociaciones observables, adaptado de Brand, [2021](#); Civelek, [2018](#).

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3.1)$$

donde:

- ρ_s : Coeficiente de correlación de Spearman.
- d_i : Diferencia entre los rangos de los pares de observaciones.
- n : Número total de observaciones.

3.2.2. Análisis de trayectorias

Los Modelos de Ecuaciones Estructurales (MES) son herramientas estadísticas multivariantes, también conocidas como análisis de estructuras de covarianza, que examinan relaciones directas (causales) e indirectas entre variables. Los MES incorporan tanto variables observables como latentes, siendo el Análisis de Trayectorias (AT) un método específicamente enfocado en las variables observables, identificando efectos directos e indirectos entre ellas (Civelek, [2018](#); Lepera, [2021](#)).

Las convenciones gráficas del modelo AT (Brand, [2021](#); Civelek, [2018](#)) se ilustran en la Figura [3.1](#), la cual representa visualmente las relaciones causales entre variables a través de asociaciones observables.

En el AT, las variables se clasifican en dos tipos (Barbeau et al., [2019](#); Civelek, [2018](#)):

- Endógenas: Afectadas por otras variables dentro del modelo.
- Exógenas: No están influenciadas por otras variables en el mismo contexto.

Tipos de efectos en un diagrama de trayectorias, clasificados según su influencia en las interacciones entre variables (Brand, 2021; Civelek, 2018):

- Efecto directo: Influencia de una variable sobre otra sin intermediarios.
- Efecto indirecto: Una variable mediadora conecta las variables independiente y dependiente.
- Efecto total: Suma de los efectos directos e indirectos.

3.2.3. Métricas estadísticas en análisis de trayectorias

Las métricas estadísticas evalúan el desempeño del modelo en relación con los datos observados, cuantificando su ajuste y capacidad predictiva. La Tabla 3.1 presenta las métricas utilizadas en el AT (Jordán-Muiños, 2021; Lepera, 2021) para evaluar si el modelo teórico representa adecuadamente la estructura subyacente de los datos.

El RMSEA y el SRMR son índices de ajuste absoluto que evalúan qué tan bien la matriz de covarianza se alinea con el modelo implícito, donde valores más bajos indican una mejor calidad de ajuste. El GFI, NFI y CFI son índices de ajuste incremental que miden las mejoras respecto a un modelo de referencia (Castillo y Ruiz, 2025; Jordán-Muiños, 2021).

Tabla 3.1. Métricas estadísticas para evaluar el ajuste del modelo en el análisis de trayectorias

Métricas	Descripción	Criterio
RMSEA	Error cuadrático medio de aproximación	<0.05
SRMR	Raíz cuadrática media normalizada residual	<0.05
GFI	Índice de bondad de ajuste	≥0.95
NFI	Índice de ajuste normalizado	≥0.95
CFI	Índice de ajuste comparativo	≥0.95

3.2.4. Prueba de Kolmogorov-Smirnov

En estadística, las pruebas paramétricas y no paramétricas se utilizan ampliamente para el análisis de datos. Una prueba no paramétrica es la de Kolmogorov-Smirnov, la cual evalúa si una muestra sigue una distribución normal.

Esta prueba considera dos hipótesis a un nivel de significancia de $p \geq 0.05$:

- H_0 : Los datos siguen una distribución normal.
- H_1 : Los datos no siguen una distribución normal.

La estadística de Kolmogorov-Smirnov, presentada en la Ecuación (3.2), mide la máxima distancia vertical entre funciones de distribución acumulada empíricas o entre una función acumulada empírica y una teórica Luzuriaga et al., 2023; Platas, 2021. La prueba determina si la hipótesis nula debe ser rechazada en función del nivel de significancia.

$$D_n = \sup_x |F_n(x) - F(x)| \quad (3.2)$$

Donde:

- D_n : Estadístico de Kolmogorov-Smirnov, que mide la distancia máxima.
- \sup_x : Supremo (el valor máximo) de las diferencias absolutas entre las funciones.
- $F_n(x)$: Función de distribución acumulada empírica basada en los datos de la muestra.
- $F(x)$: Función de distribución acumulada teórica bajo la hipótesis nula.

3.2.5. Gráficos Q-Q

Los gráficos de cuantiles normales, o gráficos Q-Q, comparan la distribución de un conjunto de datos con una distribución teórica, como la distribución normal Luzuriaga et al., 2023. En estos gráficos, los cuantiles de los datos observados se ubican en el eje horizontal, mientras que los cuantiles esperados de la distribución teórica se colocan en el eje vertical. Cuando los datos siguen la distribución teórica, los puntos se alinean en forma diagonal. Las desviaciones indican posibles discrepancias entre el conjunto de datos y el modelo teórico.

3.2.6. Transformación de datos

Las transformaciones se aplican cuando los datos no siguen una distribución normal, con el fin de mejorar la estabilidad y la interpretabilidad de los modelos. Las transformaciones aplicadas en este estudio son Ayala, [2023](#); Hamasha et al., [2022](#):

- Transformación logarítmica: Utilizada para datos positivos con sesgo a la derecha; se puede añadir una constante si existen valores cero o negativos. Se define en la Ecuación [\(3.3\)](#).

$$y = \log(x + a). \quad (3.3)$$

- Transformación de raíz cuadrada: Adecuado para datos de conteo no negativos con una asimetría moderada hacia la derecha, especialmente cuando la varianza aumenta con la media, y se representa en la Ecuación [\(3.4\)](#).

$$y = \sqrt{x}. \quad (3.4)$$

- Transformación recíproca: Ayuda a reducir el impacto de valores grandes, pero no puede aplicarse a cero, ya que el recíproco no está definido. Se expresa en la Ecuación [\(3.5\)](#).

$$y = \frac{1}{x}. \quad (3.5)$$

- Transformación Box-Cox: Un enfoque generalizado que requiere datos positivos, expresado en la Ecuación [\(3.6\)](#).

$$y = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0. \\ \log(x), & \text{si } \lambda = 0. \end{cases} \quad (3.6)$$

Donde x denota el valor original, y el valor transformado, a la constante de desplazamiento y λ un parámetro estimado a partir de los datos para mejorar la normalidad.



Figura 3.2. Etapas aplicadas durante el desarrollo del análisis.

3.3. Estudio experimental

El estudio se llevó a cabo en dos fases: la primera consistió en el preprocesamiento del conjunto de datos, y la segunda en la implementación del análisis de trayectorias para construir el modelo gráfico causal. Todos los análisis se realizaron en R versión 4.3.2 (Windows 11), utilizando las siguientes funciones y paquetes:

- Funciones estadísticas.
- Lavaan para MES/AT.
- SemPlot para la representación gráfica.

El conjunto de datos y los scripts en R están disponibles bajo solicitud.

3.3.1. Preprocesamiento de datos

Como se muestra en la Figura 3.2, el preprocesamiento de los datos involucró cinco pasos: *a)* comprender la fuente del conjunto de datos, *b)* analizar su estructura, *c)* eliminar variables irrelevantes, *d)* tratar los valores atípicos, y *e)* imputar valores faltantes para garantizar consistencia y preparación analítica.

a) Conjunto de datos: Como se muestra en la Figura 3.2, los datos fueron recolectados por un investigador de la División Académica Multidisciplinaria de Comalcalco (DAMC), Universidad Juárez Autónoma de Tabasco (UJAT). El conjunto de datos contiene información de mujeres embarazadas que participaron en campañas de embarazo saludable (agosto 2018–enero 2020) en comunidades rurales y urbanas de Tabasco, México. Incluye datos sociodemográficos, presencia bacteriana y diagnóstico de VB.

b) *Estructura de los datos*: Para comprender la estructura del conjunto de datos, se realizó un análisis exploratorio. La Tabla 3.2 resume sus principales aspectos, mostrando un desbalance entre clases, con VB- como la clase mayoritaria. Sin embargo, este desbalance no afecta la metodología de AT, ya que AT se basa en matrices de covarianza o correlación en lugar de proporciones de clases Brand, 2021; Lepera, 2021.

El conjunto de datos comprende tres clases de diagnóstico de VB:

- VB+: VB positiva, indicando presencia de la condición.
- Indeterminado: Casos sin un diagnóstico claro.
- VB-: Microbiota normal.

Tabla 3.2. Resumen de la estructura de los datos.

Aspecto	Descripción
Atributos	87 variables
Instancias	132 observaciones
Tipos de atributos	41 categóricos y 46 numéricos
Observaciones por clase	VB+ (32), Indeterminado (5) y VB- (95)
Datos faltantes	609 valores
Valores atípicos	143 valores

Además de la distribución por clases, los valores faltantes y los valores atípicos fueron examinados como parte del análisis exploratorio. Se utilizaron diagramas de caja y bigotes para identificar valores atípicos, visualizando la distribución a través de valores mínimos, primer cuartil, mediana, tercer cuartil y valores máximos Madrid et al., 2022, ver Figura 3.3.

c) *Eliminación de variables*: Posteriormente, se realizaron consultas con el equipo clínico-biológico responsable de la creación de los datos para identificar variables no relacionadas con la VB. Se eliminaron las variables asociadas al Virus del Papiloma Humano (VPH), ya que estudios previos no encontraron una asociación significativa con la VB Sanchez-Garcia et al., 2019. Este refinamiento redujo el conjunto de datos a 72 atributos y 132 instancias, conservando únicamente las variables relevantes para los objetivos del estudio.

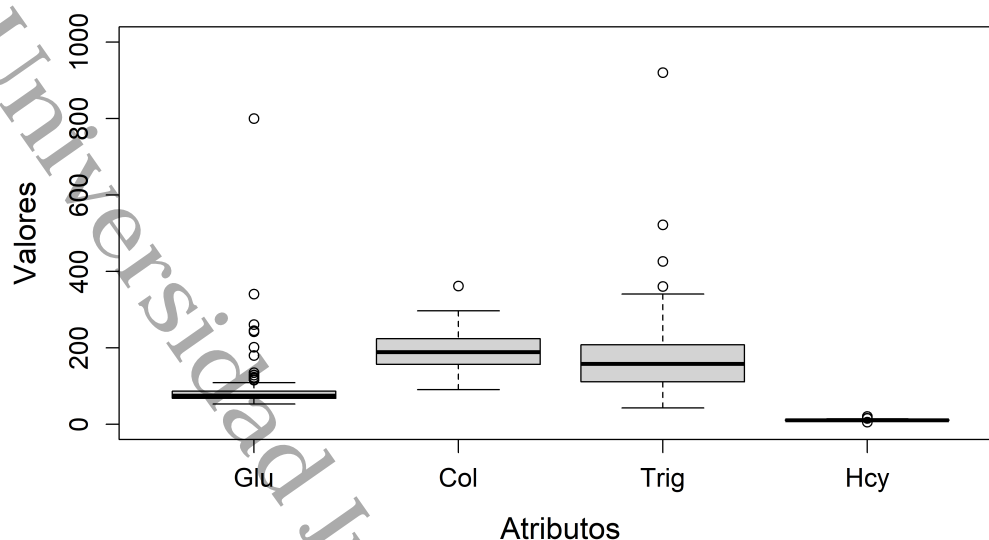


Figura 3.3. Diagramas de caja y bigotes de *Glu*, *Col*, *Trig* y *Hgb*, mostrando la distribución y variabilidad de los niveles de glucosa, colesterol, triglicéridos y hemoglobina.

d) *Tratamiento de valores atípicos:* Para garantizar la calidad y confiabilidad de los resultados, se abordaron los valores faltantes y los valores atípicos Bouza, 2023. Un análisis detallado basado en diagramas de caja identificó valores atípicos en Glucosa (*Glu*), Colesterol (*Col*), Triglicéridos (*Trig*) y Hemoglobina (*Hgb*). Estos atributos podrían corresponder a valores reales y reflejar condiciones adicionales de las pacientes; por ello, el equipo clínico-biológico recomienda conservarlos para un análisis más integral. La Figura 3.3 presenta diagramas de caja que ilustran su distribución y los valores fuera de rango preservados.

e) *Imputación de valores:* Posteriormente, los valores faltantes fueron tratados mediante imputación, reemplazando las entradas nulas para evitar que los vacíos afectaran el análisis. Para variables numéricas (por ejemplo, edad, salario), los valores faltantes se sustituyeron por la media del atributo, mientras que para variables categóricas (por ejemplo, género, color), se utilizó el valor más frecuente (moda) Alwateer et al., 2024; Rashid y Gupta, 2021.

En este estudio, la imputación se realizó con base en el diagnóstico de VB. Para variables categóricas, las instancias completas de cada atributo se agruparon por diagnóstico (VB+, Indeterminado o VB-), y la moda de cada grupo se asignó a los valores faltantes correspondientes a ese diagnóstico específico. Para variables numéricas, se aplicó la media siguiendo un proceso similar.

Las Tablas 3.3 y 3.4 muestran los datos imputados para las primeras diez observaciones de las

variables número de embarazos (*N.EMBARAZO*) y flujo vaginal (*FlujoV*). La columna *DxVBNoMh* (Diagnóstico de Vaginosis Bacteriana sin *Mycoplasma hominis*) indica el diagnóstico de VB: 1 para VB+ (positivo), 2 para Indeterminado y 3 para VB- (microbiota normal).

Tabla 3.3. Datos imputados con la media del atributo *N.EMBARAZO*.

Instancia	Valor Original	Valor Imputado	<i>DxVBNoMh</i>
1	1	1.000000	3
2	NA	2.500000	1
3	NA	2.571429	3
4	NA	2.571429	3
5	2	2.000000	3
6	NA	2.571429	3
7	1	1.000000	3
8	1	1.000000	3
9	1	1.000000	3
10	3	3.000000	3

Tabla 3.4. Datos imputados con la moda del atributo *FlujoV*.

Instancia	Valor Original	Valor Imputado	<i>DxVBNoMh</i>
1	2	2	3
2	NA	2	1
3	NA	2	3
4	NA	2	3
5	1	1	3
6	NA	2	3
7	1	1	3
8	2	2	3
9	2	2	3
10	2	2	3

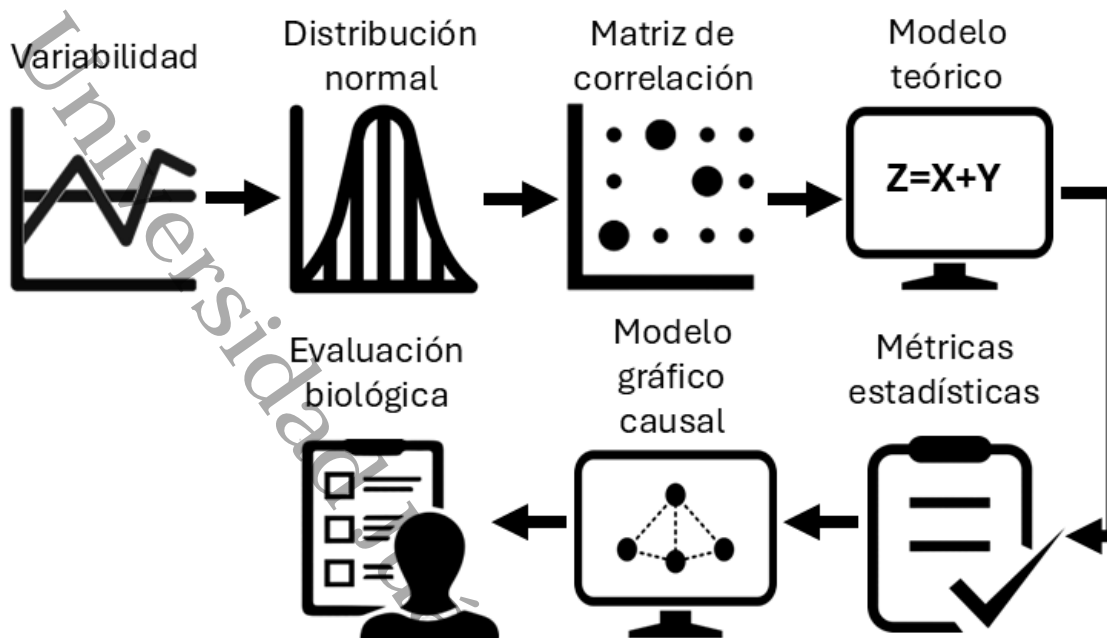


Figura 3.4. Metodología aplicada para el análisis de trayectorias: desde la exploración estadística hasta la evaluación del modelo gráfico causal.

El conjunto de datos original incluía 87 variables y 132 instancias. Después de eliminar los atributos irrelevantes, permanecieron 72 variables, con el mismo número de instancias. Los valores atípicos fueron identificados y conservados siguiendo las recomendaciones del equipo clínico-biológico, quienes los consideraron relevantes para la interpretación y el análisis. Los valores faltantes se imputaron utilizando la moda para las variables categóricas y la media para las variables numéricas, garantizando la completitud y coherencia para el análisis posterior.

3.3.2. Aplicación del análisis de trayectorias

La Figura 3.4 ilustra los pasos de la segunda fase del estudio, la cual se centró en el análisis de trayectorias. Esta etapa incluyó: *a)* evaluación de la variabilidad, *b)* análisis de la distribución, *c)* exploración de la estructura de correlación, *d)* modelado teórico, *e)* validación mediante métricas estadísticas, *f)* construcción del modelo gráfico causal y *g)* revisión por el equipo experto.

a) Variabilidad: Garantizar la variabilidad de los datos es crucial para la validez del AT, ya que la estimación de parámetros en el diagrama de trayectorias depende de la varianza y las covarianzas Castillo y Ruiz, 2025; Lepera, 2021. Por ello, se calculó la varianza para confirmar que las variables presentaban suficiente variabilidad. La Tabla 3.5 muestra los valores de varianza

de algunas variables del conjunto de datos.

Tabla 3.5. Varianza de variables ilustrativas en el conjunto de datos.

Atributo	Varianza
<i>Edad</i>	74.06
Semanas de gestación (<i>SemanasGesta</i>)	55.83
Glucosa (<i>Glu</i>)	5431.24
Colesterol (<i>Col</i>)	2335.83
<i>Gardnerella vaginalis</i> (<i>Gv</i>)	116.56
<i>Atopobium vaginae</i> (<i>Av</i>)	111.25
Bacterias asociadas a Vaginosis Bacteriana Tipo 2 (<i>BVAB2</i>)	92.58
<i>Lactobacillus crispatus</i> (<i>Lcrispatus</i>)	158.32
<i>Lactobacillus iners</i> (<i>Liners</i>)	115.07
<i>Lactobacillus jensenii</i> (<i>Ljensenii</i>)	143.58

b) *Distribución Normal*: Después de verificar la variabilidad, se analizó la distribución de las variables para determinar si la matriz de correlación de Pearson o la de Spearman sería la más adecuada para evaluar las correlaciones. El gráfico Q-Q y la prueba de normalidad de Kolmogorov-Smirnov evaluaron la adherencia a una distribución normal y se aplicaron a cada atributo del conjunto de datos, revelando que ninguno presentó normalidad.

Por ejemplo, las Figuras 3.5 y 3.6 muestran los gráficos Q-Q para semanas de gestación (*SemanaGesta*) y *Atopobium vaginae* (*Av*), ambos indicando una distribución no normal. Los resultados de la prueba mostraron $p = 0.0001741$ con $D = 0.18819$ para *SemanaGesta* y $p < 2.2 \times 10^{-16}$ con $D = 0.40474$ para *Av*, ambos con valores de p menores a 0.05, confirmando una desviación de la normalidad.

Se aplicaron diferentes técnicas de transformación, incluyendo logarítmica, raíz cuadrada, recíproca y Box-Cox, para mejorar las características de la distribución. Sin embargo, debido a la naturaleza de los datos, ninguna logró alcanzar la normalidad. La Tabla 3.6 presenta los valores p de la prueba de Kolmogorov-Smirnov aplicada a cada variable ilustrativa: semanas de gestación (*SemanasGesta*), *Mycoplasma hominis* (*Mh*), *Atopobium vaginae* (*Av*), bacterias asociadas

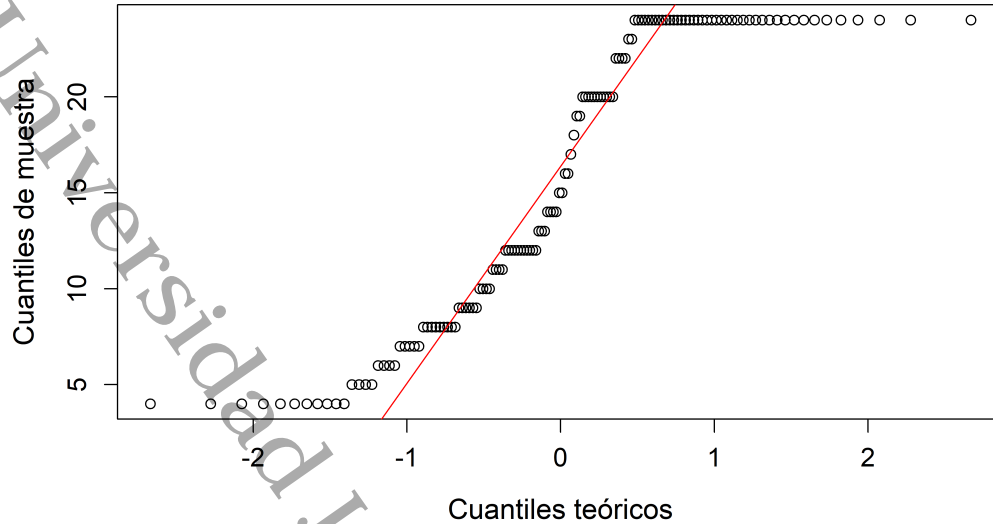


Figura 3.5. Gráficos Q-Q de *SemanasGesta*, evaluando la distribución de las semanas de gestación en comparación con una distribución normal teórica.

a vaginosis bacteriana tipo 2 (*BVAB2*) y *Lactobacillus crispatus* (*Lcrispatus*), antes y después de la transformación. Los valores N/A indican que la transformación no fue aplicable debido a la presencia de ceros.

Tabla 3.6. Prueba de Kolmogorov-Smirnov (*p*-values) antes y después de las transformaciones aplicadas a cada variable ilustrativa.

Atributo	Original	Logarítmica	Raíz cuadrada	Recíproca	Box-Cox
<i>SemanasGesta</i>	0.0001741	7.878×10^{-5}	0.0001983	6.869×10^{-6}	$< 2.2 \times 10^{-16}$
<i>Mh</i>	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
<i>Av</i>	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	N/A	N/A
<i>BVAB2</i>	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	N/A	N/A
<i>Lcrispatus</i>	1.583×10^{-7}	3.553×10^{-12}	8.09×10^{-9}	N/A	N/A

c) *Matriz de Correlación*: Dado que el conjunto de datos no sigue una distribución normal, se utilizó la matriz de correlación de Spearman para evaluar las relaciones entre las variables y su asociación con el diagnóstico de VB. Se identificaron cinco variables clave vinculadas al atributo de clase (*DxVBNoMh*): *Mycoplasma hominis* (*Mh*), *Atopobium vaginae* (*Av*), *Gardnerella vaginalis* (*Gv*), *Megasphaera Phylotype I* (*MT1*) y Bacteria Asociada a Vaginosis Bacteriana Tipo 2 (*BVAB2*) García et al., 2025.

La matriz de correlación presenta los coeficientes entre estas variables y el diagnóstico de

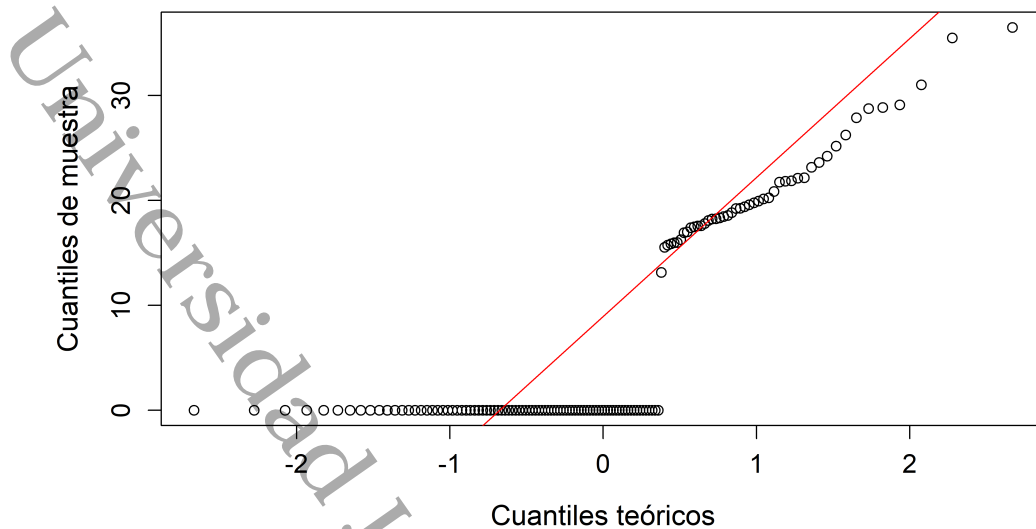


Figura 3.6. Gráficos Q-Q de A_v , evaluando la distribución de *Atopobium vaginae* en comparación con una distribución normal teórica.

VB. La Figura 3.7 incluye tres categorías diagnósticas: VB+ (presencia de la condición), Indeterminado (diagnóstico no claro) y VB- (microbiota normal), mientras que la Figura 3.8 considera únicamente VB+ y VB-. La comparación de las matrices revela que los valores de correlación son muy similares, lo que sugiere que la clase Indeterminado tiene un impacto mínimo y puede omitirse sin afectar los resultados.

Para investigar más a fondo la clase Indeterminado, futuras investigaciones realizarán un análisis exploratorio para determinar si existen patrones específicos que justifiquen su inclusión en modelos diagnósticos refinados. Por lo tanto, la Figura 3.8 constituye la matriz de principal interés en este estudio. La última columna presenta los coeficientes de correlación entre las variables y el diagnóstico de VB: $DxVBNoMh$ con A_v (-0.70), G_v (-0.41), $MT1$ (-0.89), $BVAB2$ (-0.83) y Mh (0.55). Aunque algunos valores son bajos, su relevancia clínica radica en su potencial para aportar información esencial en la toma de decisiones diagnósticas y terapéuticas Cáceres y Ponce, 2023; Poaquiza y Gavilanes, 2022.

d) *Modelo Teórico*: Con base en la matriz de correlación, se construyó un modelo teórico como requisito previo para obtener el MGC a partir del conjunto de datos. Este modelo está representado por la Ecuación (3.7), la cual especifica la estructura del modelo teórico utilizado para construir el Modelo Gráfico Causal (MGC). El símbolo \sim denota una relación direccional, donde la variable a la izquierda se modela como una función de las que aparecen a la derecha.



Figura 3.7. Matriz de correlación de *DxVBNomh* (categorías de diagnóstico de VB: VB+, Indeterminado, VB-) y presencia bacteriana, destacando sus asociaciones.

La primera línea indica que el diagnóstico de VB sin *Mycoplasma hominis* (*DxVBNomh*) se predice utilizando cinco variables bacterianas: *Mh*, *Av*, *Gv*, *MT1* y *BVAB2*. Las líneas posteriores modelan a *MT1* como una variable dependiente influenciada individualmente por cada una de las otras bacterias, reflejando su papel mediador en la vía causal. Esta estructura sugiere que *MT1* puede actuar como un conducto a través del cual otras bacterias ejercen influencia sobre el diagnóstico de VB.

ModC ←

$$'DxVBNomh \sim Mh + Av + Gv + MT1 + BVAB2$$

$$MT1 \sim Mh$$

$$MT1 \sim Av$$

$$MT1 \sim Gv$$

$$MT1 \sim BVAB2'$$

(3.7)



Figura 3.8. Matriz de correlación de DxVBNomh (categorías de diagnóstico de VB: VB+, VB-) y presencia bacteriana, destacando sus asociaciones.

e) *Métricas Estadísticas:* Después de obtener el modelo teórico, debe evaluarse su ajuste. La Tabla 3.7 presenta los valores derivados de aplicar las métricas estadísticas de la Tabla 3.1 a la Ecuación (3.7).

Tabla 3.7. Valores de las métricas del modelo.

Indicador	Valor	Valor Ideal	Buen Ajuste
RMSEA	0.00	< 0.05	Sí
SRMR	0.00	< 0.05	Sí
GFI	1.00	≥ 0.95	Sí
NFI	1.00	≥ 0.95	Sí
CFI	1.00	≥ 0.95	Sí

f) *Modelo Gráfico Causal:* Con base en el modelo teórico de la Ecuación (3.7) y las métricas estadísticas de la Tabla 3.7, se construyó el MGC (Figura 3.9) para representar visualmente las

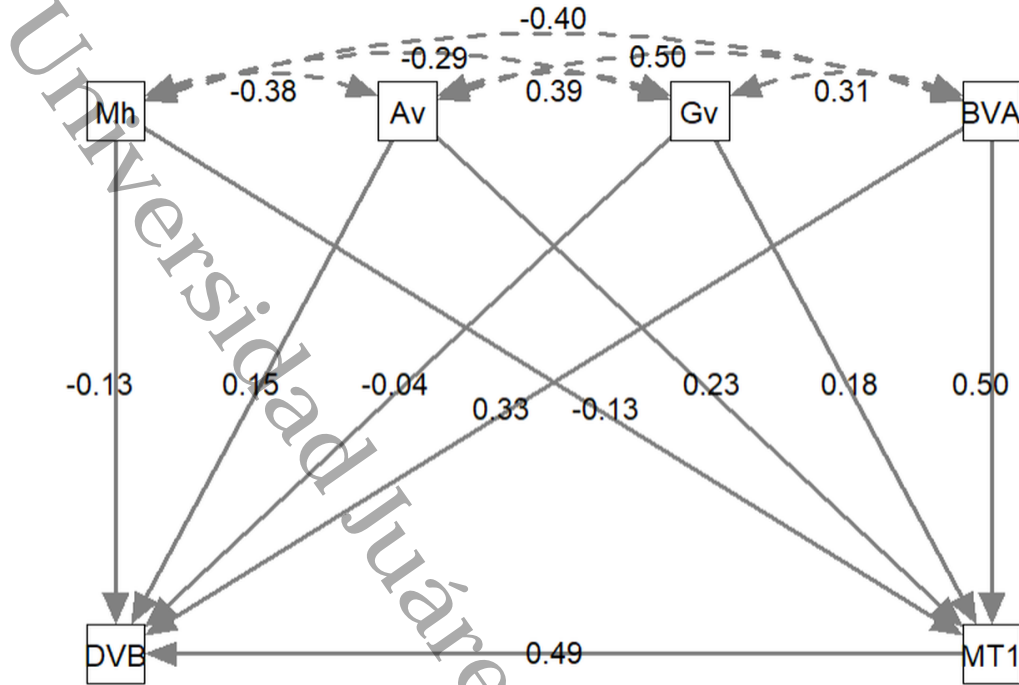


Figura 3.9. Modelo gráfico causal que ilustra las relaciones entre atributos bacterianos predictivos y el diagnóstico de VB.

relaciones causales e indirectas entre las variables analizadas y su impacto en el DVB. Debido a simplificaciones en la programación en R, algunas variables fueron abreviadas: *DxVBNoMh* como DVB y *BVAB2* como BVA. El MGC ilustra relaciones causales e indirectas (líneas sólidas) entre todas las variables y DVB, así como covariaciones (líneas punteadas). Los coeficientes de trayectoria indican la fuerza de la asociación, con valores cercanos a ± 1 representando asociaciones fuertes y valores próximos a 0 indicando asociaciones débiles.

La relación causal más fuerte observada fue entre *MT1* y DVB, con un coeficiente de 0.49, lo que sugiere que un aumento en el nivel de *MT1* se asocia con una mayor probabilidad de diagnóstico de DVB. La relación causal más débil fue entre *Gv* y DVB, con un coeficiente de -0.04, indicando una débil asociación negativa. Además, se identificaron relaciones indirectas, como la conexión entre *BVAB2* y DVB a través de *MT1*, donde *MT1* actúa como variable intermedia, con coeficientes de 0.50 y 0.49, respectivamente.

Los efectos totales se calcularon sumando los efectos directos e indirectos mediante coeficientes de trayectoria estandarizados Flora et al., 2025; Manzano, 2017. El efecto directo corresponde al coeficiente de trayectoria que vincula una variable con otra, representando su relación

inmediata. Los efectos indirectos se determinaron multiplicando los coeficientes a lo largo de las trayectorias causales. La Tabla 3.8 resume los efectos totales de las variables predictoras sobre el diagnóstico de VB.

Tabla 3.8. Efectos totales de las bacterias con $DxVBNh$ (DVB). Con base en la Figura 3.9, el efecto calculado se obtiene mediante la suma algebraica del efecto directo con el efecto indirecto, el cual a su vez se obtiene multiplicando los coeficientes de las flechas.

Atributos	Efecto Calculado	Efectos Totales en DVB
<i>Mh</i>	$-0.13 + (-0.13 \times 0.49)$	-0.1937
<i>Av</i>	$0.15 + (0.23 \times 0.49)$	0.2627
<i>Gv</i>	$-0.04 + (0.18 \times 0.49)$	0.0482
<i>MT1</i>	0.49	0.4900
<i>BVAB2</i>	$0.33 + (0.50 \times 0.49)$	0.5750

Una interpretación de la Tabla 3.8 se presenta a continuación:

- *Mh* (-0.1937): Reduce ligeramente la probabilidad de DVB, sugiriendo un modesto efecto protector.
- *Av* (0.2627): Asociado positivamente con DVB, indicando una mayor probabilidad de diagnóstico.
- *Gv* (0.0482): Correlación positiva débil; un pequeño aumento en *Gv* incrementa ligeramente la probabilidad de diagnóstico.
- *MT1* (0.4900): Impacto positivo fuerte, aumentando significativamente la probabilidad de DVB.
- *BVAB2* (0.5750): Mayor correlación positiva, indicando una fuerte asociación con el riesgo de DVB.

g) Evaluación Biológica: Finalmente, el MGC fue presentado al equipo clínico-biológico para su interpretación, validando las relaciones causales identificadas y evaluando su relevancia en el diagnóstico de VB. El análisis de trayectorias visualiza las relaciones bacterianas mediante un diagrama, ilustrando su conexión con la VB. En este estudio, las bacterias seleccionadas (*Av*,

Gv, *MT1*, *BVAB2* y *Mh*) coinciden con aquellas más comúnmente asociadas a la VB, según lo reportado en la literatura especializada en biología. El diagrama resalta las interacciones directas e indirectas, permitiendo calcular el efecto total de cada bacteria sobre el diagnóstico e identificar la más influyente.

3.4. Resultados

Este estudio analizó un conjunto de datos de mujeres embarazadas a las que se les realizó la prueba de VB, aplicando Análisis de Trayectorias (AT) para identificar bacterias coexistentes que influyen en el diagnóstico. El conjunto de datos original contenía 87 atributos y 132 instancias, categorizadas en vaginosis bacteriana, indeterminado y microbiota normal. Después del análisis exploratorio, se refinó a 72 atributos, considerando únicamente vaginosis bacteriana y microbiota normal.

Dada la distribución no normal de los datos, una matriz de correlación de Spearman identificó cinco bacterias significativamente asociadas con VB: *Mycoplasma hominis* (*Mh*), *Atopobium vaginae* (*Av*), *Gardnerella vaginalis* (*Gv*), *Megasphaera Phylotype I* (*MT1*) y Bacteria Asociada a VB Tipo 2 (*BVAB2*).

La evaluación estadística confirmó el ajuste del modelo (RMSEA = 0.00, SRMR = 0.00, GFI = 1.00, NFI = 1.00, CFI = 1.00), respaldando la construcción del Modelo Gráfico Causal (MGC). Este modelo visualizó efectos directos e indirectos entre las bacterias y el diagnóstico de VB.

Entre las bacterias identificadas, *MT1* y *BVAB2* mostraron los efectos más fuertes sobre el diagnóstico de VB, con valores de efecto total de 0.4900 y 0.5750, respectivamente. Las bacterias restantes contribuyeron de la siguiente manera: *Mh* (-0.1937), *Av* (0.2627) y *Gv* (0.0482), demostrando interacciones únicas en el ecosistema microbiológico.

La validación experta confirmó que el MGC es una herramienta visual efectiva para analizar las interacciones bacterianas y su impacto en el diagnóstico de VB, reforzando la relevancia del modelado causal en la investigación microbiológica.

3.5. Discusión

La metodología de AT identificó cinco bacterias significativamente asociadas con el diagnóstico de VB, siendo *MT1* y *BVAB2* las que mostraron los efectos más influyentes. Mientras que *Av* presentó una relación positiva de menor magnitud, *Gv* tuvo un impacto positivo débil y *Mh* mostró una posible asociación inversa.

A diferencia de los modelos de aprendizaje automático que identifican correlaciones bacterianas pero no cuantifican la influencia, PA permite evaluar los efectos directos e indirectos, proporcionando una comprensión más profunda del diagnóstico de VB. Las bacterias identificadas (*Gv*, *Av*, *Mh* y *MT1*) coinciden con hallazgos previos, reforzando la validez del MGC.

Aunque el tamaño de la muestra es relativamente pequeño (132 casos reales), AT se basa en la correcta especificación del modelo (utilizando matrices de correlación y covarianza) en lugar del balance de clases, asegurando estimaciones estables de los coeficientes. Una validación futura con conjuntos de datos más grandes fortalecerá su validez externa y mejorará su aplicabilidad clínica, demostrando el valor de AA en la investigación microbiológica.

3.6. Conclusiones

El Análisis de Trayectorias (AT) identificó de manera efectiva relaciones causales en el diagnóstico de VB, revelando cinco bacterias clave que influyen en su probabilidad. Entre ellas, *MT1* y *BVAB2* tuvieron los efectos más influyentes, mientras que *Mh* mostró una posible asociación inversa.

Estos hallazgos coinciden con la literatura previa y fueron validados por expertos clínico-biológicos, confirmando la fiabilidad de la metodología. Más allá de sus ventajas analíticas, el MGC podría ofrecer información clínicamente valiosa, apoyando evaluaciones ginecológicas y la atención prenatal, particularmente en entornos con recursos limitados.

En trabajos futuros, se propone validar este MGC utilizando conjuntos de datos más grandes y diversos, explorando su aplicabilidad en investigaciones longitudinales y multicéntricas para perfeccionar la precisión diagnóstica y las estrategias de tratamiento en entornos clínicos.

Contribución de los autores: Conceptualización, M.G. y J.C.; metodología, M.G.; software,

M.G.; validación, J.C., L.R. y E.D.; análisis formal, M.G., J.C. y L.R.; investigación, M.G.; recursos, M.G.; depuración de datos, M.G. y J.C.; redacción—preparación del borrador original, M.G.; redacción—revisión y edición, J.C. y L.R.; visualización, J.C. y L.R.; supervisión, L.R.; administración del proyecto, J.C.; adquisición de financiamiento, ninguno recibido. Todos los autores han leído y aceptado la versión publicada del manuscrito.

Financiamiento: Esta investigación no recibió financiamiento externo.

Conocimiento Informado: Se obtuvo consentimiento informado de todos los sujetos involucrados en el estudio.

Disponibilidad de los datos: Los datos presentados en este estudio están disponibles a solicitud del autor correspondiente debido a la privacidad de los datos.

Conflictos de interés: Los autores declaran no tener conflictos de interés.

Referencias del capítulo

- Álvarez-Vega, M., Quirós-Mora, L. M., & Cortés-Badilla, M. V. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, 5(8), 1-12. <https://doi.org/10.31434/rms.v5i8.557>
- Alwateer, M., Atlam, E., Abd, M., Ghoneim, O., & Gad, I. (2024). Missing Data Imputation: A Comprehensive Review. *Journal of Computer and Communications*, 12, 53-75. <https://doi.org/10.4236/jcc.2024.1211004>
- Apaza, E., Cazorla, S., Condori, C., Arpasi, F., Tumi, I., Yana, W., & Quispe, J. (2022). La Correlación de Pearson o de Spearman en caracteres físicos y textiles de la fibra de alpacas. *Revista de Investigaciones Veterinarias del Perú*, 33(3), 1-19. <https://doi.org/10.15381/rivep.v33i3.22908>
- Ayala, J. (2023). Transformaciones. <https://rpubs.com/JairoAyala/TNM>
- Barbeau, K., Boileau, K., Sarr, F., & Smith, K. (2019). Path analysis in Mplus: A tutorial using a conceptual model of psychological and behavioral antecedents of bulimic symptoms in young adults. *The Quantitative Methods for Psychology*, 15(1), 38-53. <https://doi.org/10.20982/tqmp.15.1.p038>

- Bouza, C. N. (2023). Análisis exploratorio de datos univariados para la ciencia de datos. *Red Iberoamericana de Estudios Cuantitativos Aplicados*. <https://doi.org/10.13140/RG.2.2.26038.06722>
- Brand, Y. (2021). *Modelos de ecuaciones estructurales: conceptos y aplicaciones* [Tesis doctoral, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/bitstream/handle/unal/80064/1059705148.2021.pdf>
- Cáceres, M., & Ponce, R. (2023). Dependencia emocional y su relación con el riesgo suicida en adultos jóvenes. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4. <https://doi.org/10.56712/latam.v4i1.247>
- Cameron, C., Dion, M., Broce, J., Ojo, D., Drobina, E., Louis, A., Girbert, J., Fang, R., & Parker, I. (2023). Ethnic disparity in diagnosing asymptomatic bacterial vaginosis using machine learning. *npj Digital Medicine*, 6, 1-10. <https://doi.org/10.1038/s41746-023-00953-1>
- Castillo, A., & Ruiz, A. (2025). Entre lo observable y lo latente: modelos de ecuaciones estructurales e investigación social. *Revista Reflexiones*, 104, 1-27. <https://doi.org/10.15517/rr.v104i2.58268>
- Civelek, M. E. (2018). *Essentials of Structural Equation Modeling*. Zea E-Books, University of Nebraska–Lincoln. <https://digitalcommons.unl.edu/zeabook/64>
- De la Cruz-Ruiz, F., Canul-Reich, J., Rivera-López, R., & De la Cruz-Hernández, E. (2024). Impact of data balancing a multiclass dataset before the creation of association rules to study bacterial vaginosis. *Intelligent Medicine*, 4(3), 188-199. <https://doi.org/10.1016/j.imed.2023.02.001>
- Flora, D., Crone, G., & Bell, S. (2025). Effect size interpretation in structural equation models. *Structural Equation Modeling*, 32(6), 1069-1076. <https://doi.org/10.1080/10705511.2025.2459768>
- García, M., Canul, J., Rodríguez, L., & Cruz, E. (2025). Hacia un Modelo Gráfico Causal para el Análisis de la Vaginosis Bacteriana. *Ideas en Ciencia e Ingeniería*, 3, 22-34. <https://doi.org/10.36677/rici.v3i1.25016>
- González, K., Guerra, F., Villeda, G., & López, M. (2024). Frecuencia y detección molecular de *Gardnerella vaginalis* en una institución de tercer nivel. *Revista Perinatología y Reproducción Humana*, 37, 91-98. <https://doi.org/10.24875/PER.23000020>

- Hamasha, M., Ali, H., Hamasha, S., & Ahmed, A. (2022). Ultra-fine transformation of data for normality. *Heliyon*, 8, e09370. <https://doi.org/10.1016/j.heliyon.2022.e09370>
- Hernández-Gómez, H. J., Canul-Reich, J., Hernández-Ocaña, B., & De la Cruz-Hernández, E. (2023). An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis*, 27(3), 583-611. <https://doi.org/10.3233/IDA-216488>
- Jahuey-Martinez, F. J., Herrera-Ojeda, J. B., & Paredes-Sánchez, F. A. (2022). El programa R: una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria*, 23(4). <https://doi.org/10.22201/cuaieed.16076079e.2022.23.4.4>
- Jordán-Muiños, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Revista de Investigación en Psicología Social*, 7(1), 66-71. <https://www.redalyc.org/articulo.oa?id=672371335005>
- Lepera, A. (2021). Introducción a los Modelos de Ecuaciones Estructurales y su Implementación en R mediante un ejemplo. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión Económica*, 1(1), 15-37. <https://www.economicas.uba.ar/wp-content/uploads/2016/04/Lepera-Andrea-1.pdf>
- Luzuriaga, H., Espinosa, C., Haro, A., & Ortiz, H. (2023). Histograma y distribución normal: Shapiro-Wilk y Kolmogorov Smirnov aplicado en SPSS. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4, 596-607. <https://doi.org/10.56712/latam.v4i4.1242>
- Madrid, A. E., Valenzuela-Ruiz, S. M., Batanero, C., & Garzón-Guerrero, J. A. (2022). Interpretación del diagrama de caja por estudiantes universitarios de ciencias de la actividad física y deporte. *Educación Matemática*, 34(3). <https://doi.org/10.24844/em3403.10>
- Manzano, A. (2017). Introducción a los modelos de ecuaciones estructurales. *Investigación y Educación Médica*, 7, 67-72. <https://doi.org/10.1016/j.riem.2017.11.002>
- Mendivelso, F., & Rodríguez, M. (2021). Prueba no paramétrica de correlación de Spearman. *Revista Médica Sanitas*, 24, 42-45. <https://doi.org/10.26852/01234250.578>
- Molina-Arias, M., Ochoa-Sangrador, C., & Ortega-Páez, E. (2021). Correlación, Modelos de regresión. *Evidencia Pediátrica*, 17(2), 17-25. <https://evidenciasenpediatria.es/articulo/7827/correlacion-modelos-de-regresion>

- Morelli-Martínez, I., & Gamboa-Miranda, S. (2022). Vaginosis bacteriana en el embarazo: últimos avances hasta la fecha. *Revista Médica Sinergia*, 7, e838. <https://doi.org/10.31434/rms.v7i7.838>
- Pérez-Gómez, J. F., Canul-Reich, J., & De la Cruz-Hernández, E. (2020). Combination of Rankings as a Method for Biomarker Identification of Bacterial Vaginosis. *Research in Computing Science*, 149, 915-927. <https://dblp.org/rec/journals/rcs/Perez-GomezCD20>
- Platas, V. (2021). *Contrastes de normalidad* [Tesis doctoral, Universidad de Santiago de Compostela]. <https://minerva.usc.gal/rest/api/core/bitstreams/c8efafac-12ac-446a-9443-bb3ca39fa5f0/content>
- Poaquiza, L., & Gavilanes, G. (2022). El estrés y su relación con la inteligencia emocional en docentes universitarios. *Ciencia Latina Revista Científica Multidisciplinaria*, 6, 1357-1372. https://doi.org/10.37811/cl_rcm.v6i3.2301
- Rashid, W., & Gupta, M. K. (2021). A Perspective of Missing Value Imputation Approaches. *Advances in Computational Intelligence and Communication Technology: Proceedings of CICT 2019*, 1086, 307-315. https://doi.org/10.1007/978-981-15-1275-9_25
- Sagaró-del-Campo, N. M., & Zamora-Matamoros, L. (2019). Métodos gráficos en la investigación biomédica de causalidad. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, 44(4), 1-9. <http://revzoilomarinellosld.cu/index.php/zmv/article/view/1846>
- Salas, J., Angulo, L., & Mendez, E. (2022). Vaginosis bacteriana: actualización y novedad terapéutica. *UCIMED*, 5(6), 85-93. <https://doi.org/10.34192/cienciaysalud.v5i6.387>
- Sanchez-Garcia, E. K., Contreras-Paredes, A., Martinez-Abundis, E., Garcia-Chan, D., Lizano, M., & De la Cruz-Hernandez, E. (2019). Molecular epidemiology of bacterial vaginosis and its association with genital micro-organisms in asymptomatic women. *Journal of Medical Microbiology*, 68(9), 1373-1382. <https://doi.org/10.1099/jmm.0.001044>
- Sucar, L. E. (2021). *Probabilistic Graphical Models: Principles and Applications* (2nd). Springer.

Capítulo 4

Modelos gráficos causales para la transición de una microbiota vaginal saludable a la vaginosis bacteriana en mujeres embarazadas

Modelos gráficos causales para la transición de una microbiota vaginal saludable a la vaginosis bacteriana en mujeres embarazadas

Maricela García Avalos¹; Juana Canul Reich¹; Lil María Xibai Rodríguez-Henríquez^{2,3}; Erick Natividad De la Cruz Hernández¹.

Resumen Este estudio desarrolló dos Modelos Gráficos Causales (MGC) las transiciones asociadas con la Vaginosis Bacteriana (VB) e identificar las especies bacterianas clave en cada etapa. La VB es una condición causada por un desequilibrio en la microbiota vaginal, cuya composición varía entre mujeres y a lo largo de las etapas de desarrollo. Un MGC previo identificó las bacterias más influyentes en la VB, pero no abordó los cambios entre los estados de la microbiota. En este estudio, extendimos ese enfoque para capturar dichas asociaciones. Se aplicó el Análisis de Trayectorias (AT), un método de Modelado de Ecuaciones Estructurales que considera variables observables y estima efectos directos e indirectos mediante correlaciones y covarianzas. El análisis utilizó un conjunto de datos previamente recolectado por terceros, correspondiente a 132 mujeres embarazadas (4–24 semanas de gestación) que participaron en campañas de embarazo saludable en Tabasco, México, y asociados con el diagnóstico de VB. Los modelos fueron validados mediante métricas estadísticas y revisados por un experto clínico- biológico. El primer modelo, correspondiente a la transición de microbiota normal (VB-) a un estado indeterminado (I), mostró que *Megasphaera Type 1* (MT1) realiza una contribución significativa. El segundo modelo, que describe la transición del estado indeterminado (I) a VB positiva (VB+), evidenció que *Atopobium vaginae* (Av) y *Bacterial Vaginosis Associated Bacterium Type 2* (BVAB2) realizan contribuciones significativas. Estos hallazgos destacan la importancia del estado intermedio para comprender la progresión de la disbiosis y demuestran la utilidad de los MGC para identificar asociaciones en la dinámica del microbiota.

Palabras clave: Análisis de Trayectorias, Matriz de Correlación, Métricas Estadísticas, Bacterias Anaerobia

Institución de adscripción: ¹ Universidad Juárez Autónoma de Tabasco, ² Instituto Nacional de Astrofísica, Óptica y Electrónica, ³ Secretaría de Ciencia, Humanidades, Tecnología e Innovación.

4.1. Introducción

La Vaginosis Bacteriana (VB) es una alteración común del microbioma vaginal en mujeres en edad reproductiva Das et al., 2024; Mendling et al., 2019; Sethi et al., 2025. Durante el embarazo, se asocia con diversas complicaciones ginecológicas y obstétricas, incluyendo aborto espontáneo, parto prematuro, ruptura prematura de membranas y endometritis poscesárea Mendling et al., 2019; Mina-Ortiz et al., 2024; Morelli-Martínez y Gamboa-Miranda, 2022. A nivel mundial, la VB afecta aproximadamente entre el 5 % y el 49 % de las mujeres embarazadas, con variaciones según la región y los criterios diagnósticos utilizados Morelli-Martínez y Gamboa-Miranda, 2022; Park et al., 2024; Sethi et al., 2025. Los síntomas suelen incluir flujo vaginal grisáceo, húmedo y con olor desagradable, que frecuentemente empeora después de las relaciones sexuales Mendling et al., 2019; Morelli-Martínez y Gamboa-Miranda, 2022; Sethi et al., 2025.

Diversos estudios orientados a identificar bacterias asociadas con la VB han empleado técnicas como selección de características, agrupamiento, reglas de asociación y redes bayesianas Bautista Hernández et al., 2024; De la Cruz-Ruiz et al., 2024; Hernández-Gómez et al., 2023; Pérez-Gómez, Canul-Reich y De la Cruz-Hernández, 2020. El Análisis de Trayectorias (AT), definido en Brand, 2021; Sagaró-del-Campo y Zamora-Matamoros, 2019; Sucar, 2021, fue aplicado en García-Avalos et al., 2025 a un conjunto de datos de mujeres embarazadas del estado de Tabasco, México Abundis et al., 2024. Este análisis produjo un Modelo Gráfico Causal (MGC) en el que *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera* tipo 1 (MT1) y *Bacterial Vaginosis Associated Bacterium Type 2* (BVAB2) se asociaron con la VB, identificando las especies bacterianas con mayor influencia en el diagnóstico.

Con base en García-Avalos et al., 2025 y utilizando el mismo conjunto de datos Abundis et al., 2024, este estudio examina las transiciones entre los estados diagnósticos de la VB, las cuales no habían sido abordadas previamente. Aquí, “transiciones” se refiere a comparaciones analíticas entre estados diagnósticos más que a cambios temporales. Estas transiciones no implican relaciones temporales, sino comparaciones estructuradas entre estados diagnósticos dentro de un marco transversal. En el conjunto de datos, el diagnóstico de VB se clasifica en tres grupos: microbiota normal (VB-), indeterminado (I) y vaginosis bacteriana positiva (VB+). En consecuencia, se desarrollaron dos Modelos Gráficos Causales (CGM), derivados de las ecuaciones estructura-

les: el primero captura la relación entre las bacterias con los estados VB- e I, mientras el segundo examina la asociación entre las mismas bacterias con los estados I y VB+, permitiendo identificar las especies bacterianas con mayor influencia en cada etapa de transición.

En el primer modelo (VB- a I), Av (0.0260), Gv (0.0388) y MT1 (0.3184) se asociaron con la transición hacia el estado intermedio, siendo MT1 el que realizó la contribución más significativa. En contraste, Mh y BVAB2 presentaron efectos negativos (-0.4400 y -0.0188, respectivamente), lo que indica que no promueven el cambio en esta etapa. En el segundo modelo (I a VB+), Av (0.6260) y BVAB2 (0.3800) emergieron como los agentes más influyentes en la progresión hacia VB+, con predominio de Av. Mh (0.1256) también contribuyó a la transición, MT1 (0.1164) mostró un efecto menor, y Gv (-0.0138) no resultó relevante en esta fase. Estos patrones específicos por fase muestran que la relevancia de cada agente bacteriano depende del estado, lo que respalda la necesidad de modelar explícitamente las transiciones de la microbiota para comprender la progresión hacia VB+.

El resto de este artículo se organiza de la siguiente manera: la Sección 4.2 describe los materiales y métodos utilizados. La Sección 4.3 presenta el diseño experimental, destacando la aplicación del AT para la construcción de los MGC. La Sección 4.4 detalla los resultados, enfatizando las especies bacterianas más influyentes en el diagnóstico de la VB. La Sección 4.5 discute los hallazgos en relación con estudios previos. Finalmente, la Sección 4.6 resume las principales conclusiones del estudio.

4.2. Materiales y métodos

4.2.1. Matriz de correlación de Spearman

La correlación es una medida estadística utilizada para determinar el grado de asociación lineal entre dos variables cuantitativas u ordinales. La fuerza y la dirección de esta relación se describen mediante el coeficiente de correlación, cuyo valor se encuentra dentro del intervalo [-1, +1]. Cuanto más cercano esté el coeficiente a ± 1 , más fuerte será la asociación; en cambio, valores próximos a cero indican una relación débil o inexistente, lo que refleja aleatoriedad Apaza et al., 2022; Mendivelso y Rodríguez, 2021. El coeficiente de Spearman se emplea para evaluar

relaciones monótonas entre variables ordinales organizadas en rangos o jerarquías. El coeficiente de correlación de Spearman ρ_s se define de la siguiente manera:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (4.1)$$

d_i representa la diferencia entre los rangos de cada par de observaciones y n es el número total de observaciones.

4.2.2. Conjunto de datos

Este estudio se basó en un conjunto de datos compuesto por 87 variables y 132 instancias, recolectado por un investigador en biología de la Universidad Juárez Autónoma de Tabasco durante campañas de embarazo saludable realizadas en comunidades rurales y urbanas de Tabasco entre 2018 y 2020 Abundis et al., 2024. Como parte de un estudio transversal, se obtuvo una única muestra por participante, siguiendo un protocolo de muestreo estandarizado Sanchez-Garcia et al., 2019. El conjunto de datos incluyó información sociodemográfica, presencia bacteriana, Virus del Papiloma Humano e información diagnóstica de vaginosis bacteriana (VB), categorizada como VB positiva (VB+), microbiota normal (VB-) e indeterminado (I).

El conjunto de datos fue preprocesado García-Avalos et al., 2025. Las variables no relacionadas con la VB fueron excluidas en consulta con un experto clínico-biológico, asegurando que solo se conservaran los atributos relevantes para el objetivo del estudio. Esta reducción dio como resultado 72 atributos y 132 observaciones. Los valores atípicos Bouza, 2023 fueron identificados mediante diagramas de caja (boxplots); variables como glucosa y colesterol presentaron valores atípicos. Estas observaciones se conservaron de acuerdo con recomendaciones clínicas, ya que pueden reflejar condiciones adicionales del paciente y no errores de medición. Los valores faltantes se imputaron reemplazando los atributos numéricos por su media y los atributos categóricos por su moda Alwateer et al., 2024; Rashid y Gupta, 2021. La imputación se realizó mediante un enfoque estratificado basado en el diagnóstico de VB.

Las pruebas de normalidad (gráficas Q-Q y prueba de Kolmogorov-Smirnov Luzuriaga et al., 2023; Platas, 2021) indicaron que ninguna de las variables cumplía el supuesto de normalidad. En consecuencia, se aplicó la matriz de correlación de Spearman en las tres clases diagnósticas

(VB+, I y VB-) para identificar las variables más fuertemente asociadas con el diagnóstico de VB. Este análisis reveló cinco especies bacterianas clave: Mh, Av, Gv, MT1 y BVAB2. El conjunto de datos final se redujo a seis variables (cinco especies bacterianas y el diagnóstico de VB), manteniendo las 132 instancias originales García-Avalos et al., 2025.

4.2.3. Tipos de bacterias asociadas con la VB

En la VB predominan bacterias anaerobias, incluyendo tanto anaerobias facultativas como anaerobias estrictas, aunque también pueden estar presentes otras especies con diferentes requerimientos de oxígeno. Las anaerobias facultativas pueden crecer en presencia o ausencia de oxígeno, mientras que las anaerobias estrictas requieren ambientes libres de oxígeno para su crecimiento Tortora et al., 2021. En particular, *Mycoplasma hominis* (Mh) y *Gardnerella vaginalis* (Gv) son anaerobias facultativas, mientras que *Atopobium vaginae* (Av), *Megasphaera Type 1* (MT1) y *Bacterial Vaginosis Associated Bacterium Type 2* (BVAB2) son anaerobias estrictas.

4.2.4. Análisis de trayectorias

El Modelado de Ecuaciones Estructurales (MEE), también conocido como análisis de estructuras de covarianza, permite evaluar relaciones directas (causales) e indirectas entre variables. Dentro de este marco, el AT se enfoca exclusivamente en variables observables, proporcionando una representación gráfica de sus interacciones Brand, 2021; Civelek, 2018; Lepera, 2021. En este estudio, el término “causal” se refiere a relaciones direccionales especificadas dentro del marco de MEE/AT con base en los datos observados y la estructura del modelo propuesta. Estas relaciones deben interpretarse como asociaciones basadas en el modelo y no como efectos causales de intervención, ya que las explicaciones causales en estudios observacionales dependen de la especificación del modelo y no surgen directamente de los datos Cortés, 2018.

La representación gráfica del AT se estructura de la siguiente manera: los cuadrados o rectángulos representan variables observables; las flechas unidireccionales indican un efecto directo de una variable sobre otra; las flechas curvas de doble punta denotan una asociación o covarianza entre variables; y los parámetros del modelo se muestran a lo largo de las flechas correspondientes Brand, 2021; Civelek, 2018.

En este enfoque, las variables se clasifican como endógenas, es decir, aquellas que dependen de otras variables dentro del modelo, y exógenas, es decir, aquellas que no son influenciadas por ninguna otra variable Barbeau et al., 2019; Civelek, 2018. Los efectos pueden categorizarse como directos, indirectos (mediados a través de otra variable) o totales (la suma de ambos) Brand, 2021; Civelek, 2018; Flora et al., 2025; Manzano, 2017.

4.2.5. Métricas estadísticas en el análisis de trayectorias

Las métricas estadísticas permiten estimar el ajuste del modelo y su capacidad predictiva en relación con los datos observados. La Tabla 4.1 presenta los indicadores utilizados en el AT Jordán-Muiños, 2021; Lepera, 2021 para verificar la representación de la estructura subyacente.

El RMSEA y el SRMR son índices absolutos que evalúan la correspondencia entre la matriz de covarianzas observada y el modelo, donde valores más bajos reflejan un mejor ajuste. En contraste, el GFI, NFI y CFI son índices incrementales que comparan el desempeño del modelo con respecto a un modelo de referencia Castillo y Ruiz, 2025; Jordán-Muiños, 2021.

Tabla 4.1. Métricas estadísticas utilizadas para evaluar el ajuste del modelo en el análisis de trayectorias.

Métrica	Descripción	Valor esperado
RMSEA	Error cuadrático medio de aproximación	< 0.05
SRMR	Residuo cuadrático medio estandarizado	< 0.05
GFI	Índice de bondad de ajuste	≥ 0.95
NFI	Índice de ajuste normado	≥ 0.95
CFI	Índice de ajuste comparativo	≥ 0.95

4.2.6. Clasificación de los efectos totales en el análisis de trayectorias

En el análisis de las trayectorias entre variables observables, la magnitud de los efectos totales se clasificó de acuerdo con los rangos propuestos por Hair et al. Hair et al., 2019, considerando el valor absoluto de los coeficientes. Los rangos utilizados fueron los siguientes (véase la Tabla 4.2).

Tabla 4.2. Clasificación de la magnitud de los efectos totales según Hair et al. Hair et al., 2019.

Valor absoluto del coeficiente	Clasificación
≥ 0.50	Muy fuerte
0.30 – 0.49	Fuerte
0.10 – 0.29	Moderado
< 0.10	Muy débil

4.3. Estudio experimental

El estudio se llevó a cabo siguiendo las etapas ilustradas en la Fig. 4.1: (a) datos seleccionados, (b) etapas de diagnóstico, (c) matrices de correlación, (d) modelos teóricos, (e) métricas estadísticas, (f) modelos gráficos causales y (g) evaluación por un biólogo. Todos los procedimientos se implementaron en R (v. 4.3.2, Windows 11) utilizando funciones estadísticas y los paquetes lavaan (para el análisis de trayectorias) y semPlot (para la visualización estructural).

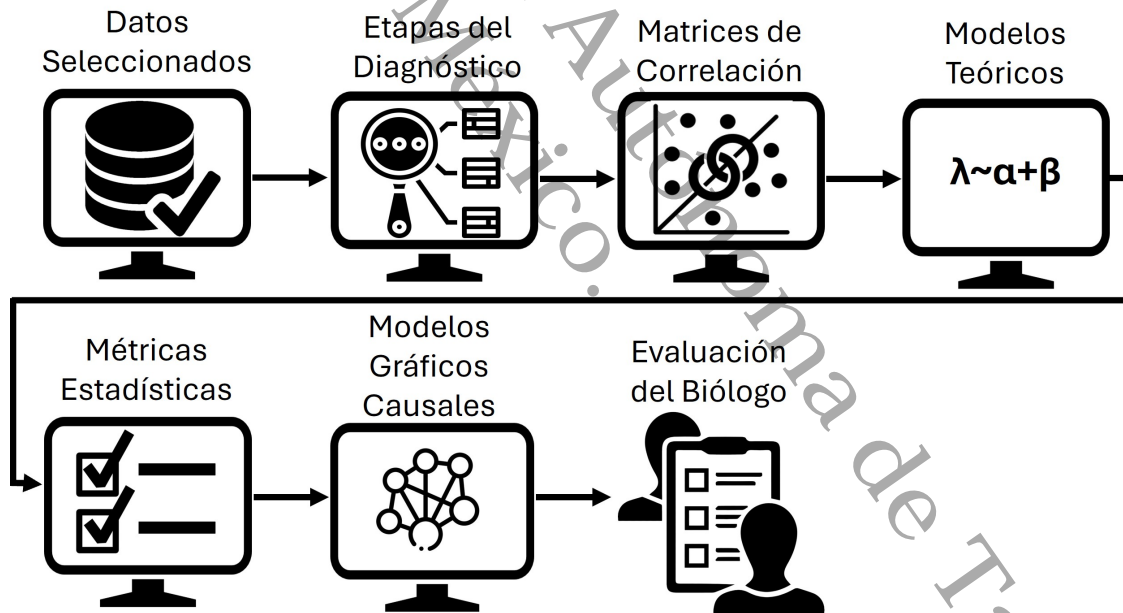


Figura 4.1. Metodología del estudio

(a) Datos seleccionados: El conjunto de datos estuvo conformado por las bacterias Mh, Av, Gv, MT1 y BVAB2, junto con el diagnóstico de Vaginosis Bacteriana sin *Mycoplasma hominis* (DxVBNoMh), distribuidos en 132 instancias. El diagnóstico clínico se clasificó en tres categorías: microbiota normal (VB–), indeterminado (I) y vaginosis bacteriana positiva (VB+).

(b) Etapas del diagnóstico: Para explorar los patrones de transición hacia un diagnóstico positivo de vaginosis bacteriana (VB+) e identificar qué perfiles bacterianos ejercen mayor influencia en la progresión hacia VB+, se realizó un análisis de trayectorias agrupando el diagnóstico en dos etapas secuenciales. La primera etapa incluyó los casos clasificados como VB- e I, mientras que la segunda comprendió los casos I y VB+. Esta segmentación permitió modelar el comportamiento transicional de las variables bacterianas, considerando el estado indeterminado como un punto intermedio entre la microbiota normal y la vaginosis bacteriana confirmada.

(c) Matrices de correlación: Como paso preliminar al modelado de trayectorias, se calcularon matrices de correlación de Spearman entre las especies bacterianas y el diagnóstico en cada etapa, como se describió anteriormente, con el fin de identificar los perfiles microbianos más asociados con cada transición diagnóstica. La Fig. 4.2 muestra la correlación entre las bacterias y el diagnóstico VB- e I, donde Mh presenta la asociación más alta (0.48). En la Fig. 4.3, correspondiente al diagnóstico I y VB+, BVAB2 destaca como la bacteria más relacionada con el diagnóstico (-0.49).

(d) Modelos teóricos: Con base en las matrices de correlación mostradas en las Figuras 4.2 y 4.3, se establecieron los modelos teóricos definidos por las ecuaciones (4.2) y (4.3). En el modelo (4.2), el símbolo \sim denota una relación direccional, donde la variable del lado izquierdo es explicada por las variables del lado derecho. La primera línea especifica que el diagnóstico de vaginosis bacteriana sin la presencia de *Mycoplasma hominis* (DxVBNoMh) se estima a partir de cinco variables bacterianas: Mh, Av, Gv, MT1 y BVAB2.

En las líneas siguientes, Mh se modela como una variable dependiente influenciada individualmente por cada una de las otras bacterias, lo que indica su posible papel como variable intermediaria. De este modo, Mh podría funcionar como un nodo de transmisión a través del cual otras bacterias ejercen su influencia sobre el diagnóstico de VB.

De manera similar, se derivó el modelo (4.3), en el cual BVAB2 aparece como variable dependiente.

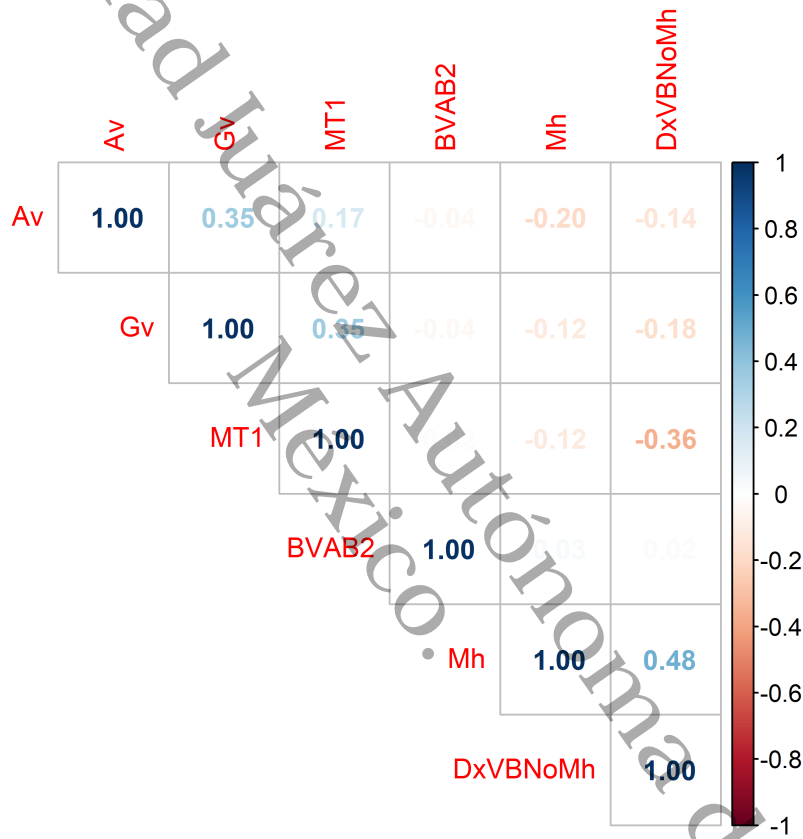


Figura 4.2. Matriz de correlación entre DxVBNomh (categorías diagnósticas: VB- e I) y la presencia bacteriana, destacando sus asociaciones.

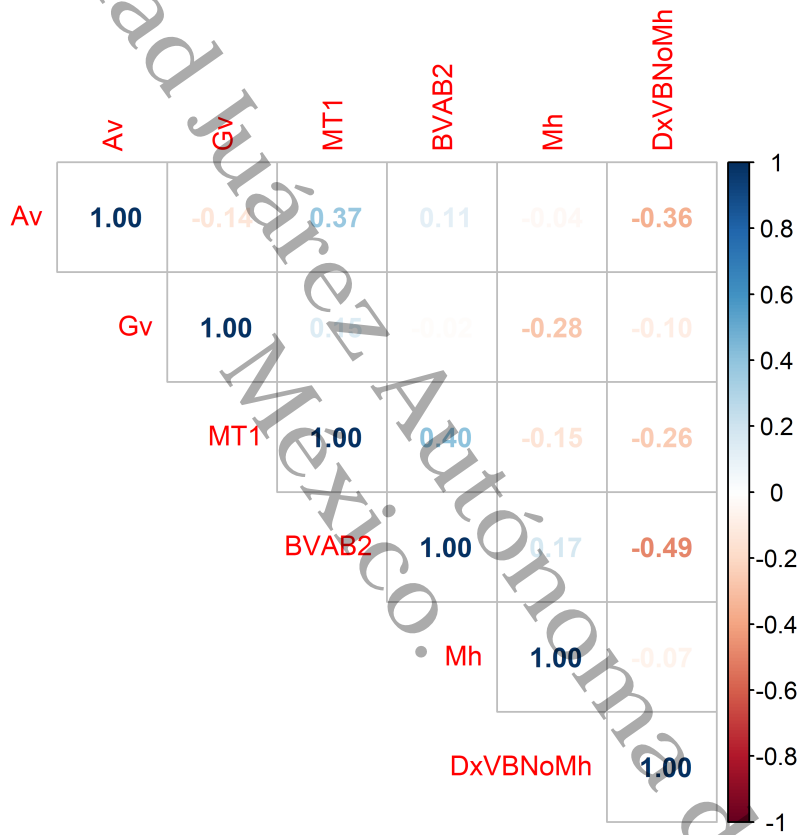


Figura 4.3. Matriz de correlación entre DxVBNomh (categorías diagnósticas: I y VB+) y la presencia bacteriana, destacando sus asociaciones.

Mod1 ←

$$'DxVBN0Mh \sim Mh + Av + Gv + MT1 + BVAB2$$

$$Mh \sim Av$$

$$Mh \sim Gv$$

$$Mh \sim MT1$$

$$Mh \sim BVAB2'$$

(4.2)

Mod2 ←

$$'DxVBN0Mh \sim Mh + Av + Gv + MT1 + BVAB2$$

$$BVAB2 \sim Mh$$

$$BVAB2 \sim Av$$

$$BVAB2 \sim Gv$$

$$BVAB2 \sim MT1'$$

(4.3)

(e) Métricas estadísticas: El modelo (4.2) y el modelo (4.3) fueron evaluados utilizando las métricas estadísticas mostradas en la Tabla 4.1. Los resultados para ambos modelos se presentan en la Tabla 4.3, donde se observaron valores idénticos. Estos resultados confirman que los modelos teóricos propuestos cumplen con los criterios establecidos de bondad de ajuste, demostrando consistencia en las métricas evaluadas.

Tabla 4.3. Valores de las métricas para el modelo (4.2) y el modelo (4.3).

Indicador	Valor	Valor ideal	Buen ajuste
RMSEA	0.00	< 0.05	Sí
SRMR	0.00	< 0.05	Sí
GFI	1.00	≥ 0.95	Sí
NFI	1.00	≥ 0.95	Sí
CFI	1.00	≥ 0.95	Sí

(f) Modelos Gráficos Causales: Posteriormente, los modelos gráficos causales (Fig. 4.4 y Fig. 4.5) fueron construidos con base en el modelo (4.2) y el modelo (4.3), ilustrando las rela-

ciones directas e indirectas entre las bacterias y el diagnóstico, lo que permitió identificar aquellas que ejercen mayor influencia en cada transición. Para simplificar la implementación en R, algunos nombres de variables fueron abreviados: DxVBNoMh como DVB y BVAB2 como BVA.

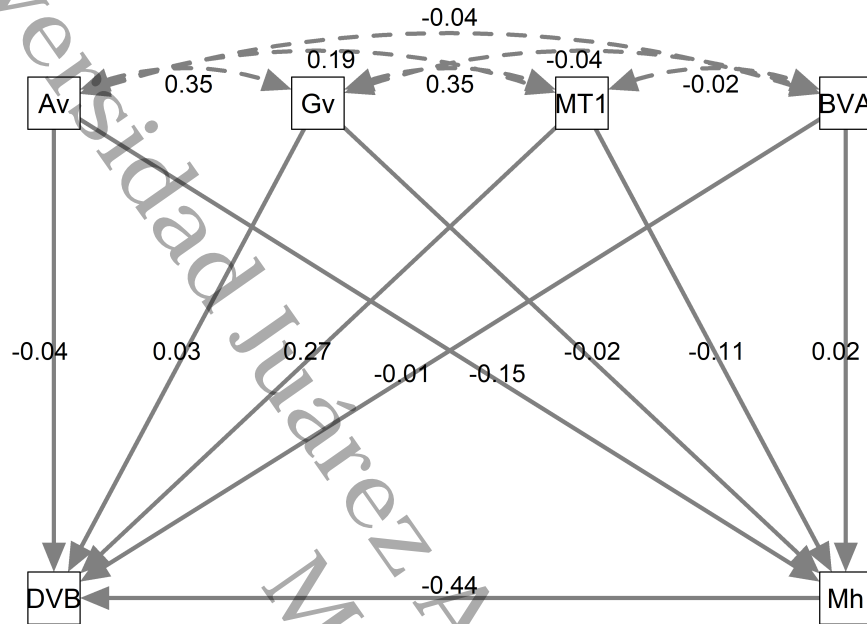


Figura 4.4. Modelo gráfico causal que representa los vínculos entre los atributos bacterianos predictivos y el diagnóstico en las categorías VB- e I. Las líneas continuas denotan efectos directos/indirectos; las líneas discontinuas denotan covariaciones.

El efecto total se calculó como la suma algebraica de los efectos directos e indirectos, obteniéndose estos últimos mediante la multiplicación de los coeficientes asociados con las flechas de conexión. En el diagrama de trayectorias (Fig. 4.4), Av ejerce un efecto directo sobre DVB de -0.04 y un efecto indirecto mediado por Mh. El efecto indirecto se calculó como $-0.15 \cdot -0.44 = 0.066$, resultando en un efecto total de $-0.04 + 0.066 = 0.0260$. Esto refleja tanto una contribución directa como una influencia mediada a través de Mh.

De manera similar, en el diagrama de trayectorias (Fig. 4.5), Mh ejerce un efecto directo sobre DVB de 0.08 y un efecto indirecto mediado por BVA. El efecto indirecto se calculó como $0.12 \cdot 0.38 = 0.0456$, obteniéndose un efecto total de $0.08 + 0.0456 = 0.1256$. Esto refleja la combinación de efectos directos y mediados a través de BVA. El mismo procedimiento se aplicó a las demás variables, permitiendo una evaluación exhaustiva de sus contribuciones individuales y mediadas a DVB. En la Tabla 4.4 se presentan los efectos totales derivados de la Fig. 4.4 y la Fig. 4.5.

Las Tablas 4.5 y 4.6 presentan la clasificación de los efectos totales y su interpretación biológica.

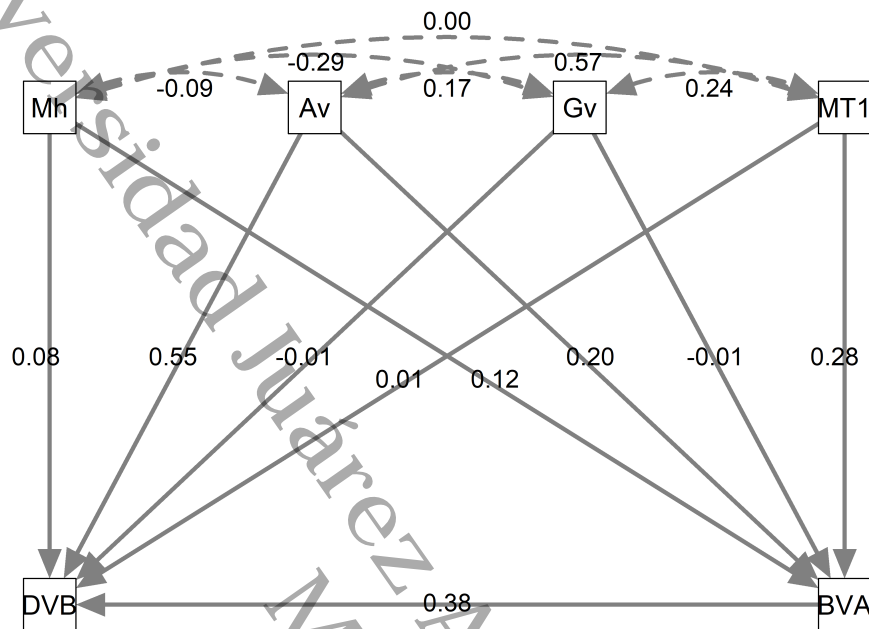


Figura 4.5. Modelo gráfico causal que representa los vínculos entre los atributos bacterianos predictivos y el diagnóstico en las categorías I y VB+. Las líneas continuas denotan efectos directos/indirectos; las líneas discontinuas denotan covariaciones.

Tabla 4.4. Efectos totales de las bacterias sobre DxVBNomh (DVB) en cada etapa.

Atributos	VB- e I	I y VB+
<i>Atopobium vaginae</i>	0.0260	0.6260
<i>Mycoplasma hominis</i>	-0.4400	0.1256
<i>Gardnerella vaginalis</i>	0.0388	-0.0138
<i>Megasphaera Type 1</i>	0.3184	0.1164
<i>Bacteria Associated with Bacterial Vaginosis Type 2</i>	-0.0188	0.3800

ca. La clasificación se realizó siguiendo los rangos propuestos por Hair et al. Hair et al., 2019, considerando la magnitud absoluta de los coeficientes. La dirección de cada efecto (positivo o negativo) fue analizada en términos de su impacto clínico sobre el diagnóstico de VB, permitiendo integrar la interpretación estadística con la relevancia biológica observada en cada etapa del modelo.

Tabla 4.5. Interpretación de los efectos totales de cada bacteria sobre el diagnóstico VB- e Indeterminado.

Bacteria (Coeficiente Total)	Impacto en el diagnóstico	Interpretación contextual
Av (0.0260)	Muy débil	Av comienza a mostrar una influencia muy ligera, posiblemente actuando como organismo acompañante en las etapas iniciales del desequilibrio.
Mh (-0.4400)	Moderado	Su efecto indica que su presencia no contribuye a la transición hacia el estado indeterminado e incluso podría retrasarla.
Gv (0.0388)	Muy débil	Gv tiene un impacto mínimo en esta transición inicial, lo que indica que su presencia por sí sola es insuficiente para alterar significativamente el estado vaginal.
MT1 (0.3184)	Moderado	MT1 contribuye en mayor medida al cambio de un estado saludable a uno indeterminado, lo que indica una posible asociación con alteraciones en la microbiota vaginal.
BVAB2 (-0.0188)	Muy débil	Su efecto indica que aún no está involucrada en el cambio hacia la disbiosis.

Capítulo 4. Modelos gráficos causales para la transición de una microbiota vaginal saludable a la vaginosis bacteriana en mujeres embarazadas

Tabla 4.6. Interpretación de los efectos totales de cada bacteria en el diagnóstico Indeterminado y VB+.

Bacteria (Coeficiente Total)	Impacto en el diagnóstico	Interpretación contextual
Av (0.6260)	Fuerte	Av presenta el efecto más fuerte en esta etapa, reforzando su papel como marcador determinante asociado al diagnóstico positivo.
Mh (0.1256)	Débil	Mh cambia de signo respecto a la etapa previa, lo que indica que puede involucrarse en fases más avanzadas del desequilibrio vaginal.
Gv (-0.0138)	Muy débil	Gv no contribuye al cambio del estado indeterminado hacia un diagnóstico positivo.
MT1 (0.1164)	Débil	MT1 continúa contribuyendo, pero con menor peso. Participa en la progresión, aunque su impacto disminuye en comparación con la etapa anterior.
BVAB2 (0.3800)	Moderado	BVAB2 está fuertemente asociada con la progresión hacia VB+, confirmando su papel como indicador predominante de disbiosis vaginal.

(g) Evaluación por un biólogo: Los MGC (Fig. 4.4 y Fig. 4.5) fueron presentados a un especialista clínico-biológico para la interpretación y validación de las relaciones identificadas en las transiciones entre microbiota normal (VB-), estado indeterminado (I) y vaginosis bacteriana positiva (VB+), incluyendo la identificación de las bacterias más influyentes en cada etapa. El experto evaluó las relaciones considerando su plausibilidad biológica, coherencia clínica y concordancia con la literatura científica.

El análisis de ambas etapas revela una secuencia microbiológica progresiva hacia la disbiosis característica de la vaginosis bacteriana (VB). En la fase temprana (VB- a I), MT1 es una bacteria que contribuye al proceso de transición, mientras que Mh no muestra una participación relevante. En la etapa avanzada (I a VB+), BVAB2 y Av emergen como marcadores predominantes de la transición hacia un estado clínico VB+, consolidando el perfil disbiótico. En este punto, Mh modifica su influencia al colaborar con otras especies en el desarrollo de la disbiosis establecida. Gv presenta un papel dinámico y variable que depende del estado del perfil microbiológico.

4.4. Resultados

Se analizaron datos de mujeres embarazadas que participaron en una campaña de embarazo saludable en Tabasco (2018–2020) y que fueron evaluadas para VB. El conjunto de datos utilizado en el estudio comprende 132 instancias e incluye cinco especies bacterianas: *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera Type 1* (MT1) y *Bacterial Vaginosis-Associated Bacterium Type 2* (BVAB2), todas asociadas con el diagnóstico DxVBNoMh, el cual se clasifica en tres categorías: microbiota normal (VB–), indeterminado (I) y vaginosis bacteriana positiva (VB+).

Para analizar el comportamiento bacteriano en relación con la progresión diagnóstica hacia VB+, se consideraron dos etapas diagnósticas: de VB– a I, y de I a VB+. Se aplicó análisis de trayectorias para construir modelos teóricos, los cuales fueron validados mediante métricas estadísticas (RMSEA = 0.00, SRMR = 0.00, GFI = 1.00, NFI = 1.00, CFI = 1.00), confirmando un ajuste óptimo. Con base en estos modelos, se generaron modelos gráficos causales (MGCs), permitiendo la visualización de relaciones directas e indirectas entre las bacterias y el diagnóstico.

Los efectos totales de cada especie bacteriana sobre los diagnósticos se estimaron a partir de los MGC, mostrando variaciones entre las etapas de transición. Durante el cambio de VB– a I, el grupo de contribuyentes positivos incluyó MT1 (moderado, 0.3184), Av (débil, 0.0260) y Gv (ligeramente positivo, 0.0388), lo que sugiere una participación temprana en el desequilibrio vaginal. En contraste, Mh mostró un efecto negativo notable (-0.4400) y BVAB2 un efecto negativo débil (-0.0188), lo que indica que no contribuyen a los cambios en esta etapa.

En la transición de I a VB+, Mh adquirió un efecto positivo (0.1256), integrándose al proceso de disbiosis, mientras que MT1 permaneció positivamente asociada (0.1164). Av mostró un efecto positivo fuerte (0.6260), seguida por BVAB2 (0.3800), ambas vinculadas con la progresión hacia una disbiosis avanzada. Gv presentó un efecto negativo leve (-0.0138), lo que indica una participación menos consistente. Estos resultados destacan perfiles bacterianos con potencial valor predictivo para el diagnóstico de VB. La validación clínica confirmó que los MGC visualizaron eficazmente las interacciones bacterianas durante la transición hacia VB+, apoyando el uso de enfoques causales basados en modelos en la investigación microbiológica.

El AT a lo largo de las etapas diagnósticas permitió visualizar las asociaciones y la magnitud

de los efectos bacterianos en la progresión de la VB. Av y BVAB2 emergieron como los principales determinantes de la transición hacia VB+, con Av actuando como el factor decisivo. Estos hallazgos refuerzan el valor predictivo de los perfiles bacterianos y destacan la utilidad de los MGC en la investigación microbiológica.

4.5. Discusión

El presente estudio aporta nuevos conocimientos sobre los patrones microbianos asociados con la VB en mujeres embarazadas, una población de particular relevancia clínica debido a su susceptibilidad a complicaciones obstétricas. Mediante la aplicación del análisis de trayectorias para generar MGCs, fue posible visualizar asociaciones basadas en modelos y cuantificar las relaciones de especies bacterianas específicas a través de las transiciones diagnósticas, aclarando así sus roles en la progresión hacia VB+.

Durante la transición de VB- a I, MT1, Av y Gv emergieron como contribuyentes positivos tempranos, lo que sugiere su participación en las etapas iniciales del desequilibrio vaginal. En contraste, Mh y BVAB2 mostraron efectos negativos, indicando ausencia de participación en esta fase temprana. Estos hallazgos resaltan que la disbiosis no comienza de manera uniforme entre las especies, sino a través de contribuciones selectivas que desestabilizan la microbiota.

En la transición posterior de I a VB+, la dinámica cambió sustancialmente. Mh pasó de un efecto negativo a uno positivo, lo que indica su participación en la progresión hacia VB+, mientras que MT1 mantuvo una asociación positiva moderada. Av y BVAB2 mostraron efectos positivos fuertes, consolidando su papel como impulsores clave de la disbiosis avanzada y del diagnóstico VB+. Sin embargo, Gv presentó un efecto negativo débil, lo que refleja que no contribuye de manera consistente. Esta progresión subraya la importancia de Av y BVAB2 como determinantes centrales de VB+, con Av actuando como el factor decisivo. Estos hallazgos son consistentes con evidencia previa que señala a BVAB2 y Av como marcadores clave del desequilibrio de la microbiota vaginal Castro-Castrillo y Duarte-Artavia, [2025](#); García-Avalos et al., [2025](#); Mendling et al., [2019](#).

Estos resultados coinciden con estudios previos que han vinculado Av y BVAB2 con estados clínicos de VB, reforzando su valor predictivo en contextos diagnósticos. Además, la capacidad

de los MGC para capturar tanto interacciones bacterianas directas como indirectas demuestra su utilidad como herramientas visuales y analíticas para comprender procesos microbianos. La integración de la validación estadística con el modelado causal fortalece la confianza en estos hallazgos y respalda la aplicación de estos enfoques en la investigación microbiológica.

En conjunto, este estudio amplía la comprensión de cómo especies bacterianas específicas contribuyen al inicio y progresión de la VB, destacando posibles biomarcadores para el diagnóstico y ofreciendo un marco metodológico para futuras investigaciones sobre disbiosis microbiana.

4.6. Conclusiones

Este estudio identificó los roles diferenciales de cinco especies bacterianas en la progresión hacia la VB en mujeres embarazadas. Los modelos gráficos causales (MGC), validados estadísticamente, revelaron que la transición de microbiota normal al estado indeterminado se caracterizó por las contribuciones de MT1 (0.3184), Av (0.0260) y Gv (0.0388).

En la transición posterior hacia VB+, Mh adquirió un efecto positivo (0.1256), MT1 permaneció positivamente asociada (0.1164), mientras que Av (0.6260) y BVAB2 (0.3800) se consolidaron como los principales determinantes, con Av actuando como el factor decisivo en el diagnóstico avanzado.

Estos hallazgos son consistentes con estudios previos y refuerzan el valor predictivo de Av y BVAB2 como biomarcadores clínicos de VB en mujeres embarazadas. Asimismo, confirman la utilidad del análisis de trayectorias para generar herramientas basadas en MGC que permiten visualizar las interacciones bacterianas y comprender los procesos de disbiosis. De manera importante, los resultados fueron evaluados y validados por un biólogo clínico, lo que fortalece su relevancia diagnóstica y microbiológica.

Como trabajo futuro, se plantea ampliar el conjunto de datos para incluir poblaciones más grandes y diversas, lo que permitirá validar estos hallazgos en distintos contextos clínicos. Además, se requieren análisis longitudinales para esclarecer la dinámica temporal de Av y BVAB2 durante el embarazo y evaluar su posible asociación con desenlaces obstétricos.

Conceptualización, M.G. y J.C.; **metodología**, M.G.; software, M.G.; **validación**, J.C., L.R. y E.D.; **análisis formal**, M.G., J.C. y L.R.; **investigación**, M.G.; **recursos**, M.G.; **curación de da-**

tos, M.G. y J.C.; **redacción—preparación del borrador original**, M.G.; **redacción—revisión y edición**, J.C. y L.R.; **visualización**, J.C. y L.R.; **supervisión**, L.R.; **administración del proyecto**, J.C.; **adquisición de financiamiento**, no se recibió financiamiento específico. Todos los autores han leído y aceptado la versión publicada del manuscrito.

Financiamiento: Esta investigación fue financiada parcialmente por la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) a través de la beca número 842958.

Comité de Ética Institucional: La revisión y aprobación ética fueron exentas para este estudio debido al uso de datos secundarios proporcionados por DAMC-UJAT, sin participación directa de sujetos humanos por parte de los autores.

Consentimiento informado: No se requirió consentimiento informado para la participación de acuerdo con la política institucional, ya que los autores no recopilaron datos directamente de sujetos humanos.

Disponibilidad de los datos: Los datos presentados en este estudio están disponibles a solicitud del autor de correspondencia debido a consideraciones de privacidad de los datos.

Conflictos de interés: Los autores declaran no tener conflictos de interés.

Las siguientes abreviaturas se utilizan en este manuscrito:

VB	Vaginosis Bacteriana
VB-	Microbiota Normal
VB+	Vaginosis Bacteriana Positiva
Mh	Mycoplasma hominis
Av	Atopobium vaginae
MT1	Megasphaera Tipo I
Gv	Gardnerella vaginalis
BVAB2	Bacteria Asociada con Vaginosis Bacteriana Tipo 2
DxVBNoMh	Diagnóstico de Vaginosis Bacteriana sin la presencia de Mycoplasma hominis
RMSEA	Raíz del Error Cuadrático Medio de Aproximación
SRMR	Residuo Cuadrático Medio Estandarizado
GFI	Índice de Bondad de Ajuste
NFI	Índice de Ajuste Normalizado
CFI	Índice de Ajuste Comparativo
PA	Análisis de Trayectorias
SEM	Modelos de Ecuaciones Estructurales
MGC	Modelos Gráficos Causales
UJAT	Universidad Juárez Autónoma de Tabasco

Referencias del capítulo

- Abundis, E. M., Hernandez-Landero, F., Escobar-Calderon, G., Gomez-Crisostomo, N., Contreras-Paredes, A., & de la Cruz-Hernandez, E. (2024). Gene expression of cardiovascular risk markers in mononuclear cells of pregnant women in relation to plasma leptin and homocysteine levels: A cross-sectional study. *International Journal of Gynaecology and Obstetrics*, 165, 350-360. <https://doi.org/10.1002/ijgo.15302>
- Alwateer, M., Atlam, E., Abd, M., Ghoneim, O., & Gad, I. (2024). Missing Data Imputation: A Comprehensive Review. *Journal of Computer and Communications*, 12, 53-75. <https://doi.org/10.4236/jcc.2024.1211004>

- Apaza, E., Cazorla, S., Condori, C., Arpasi, F., Tumi, I., Yana, W., & Quispe, J. (2022). La Corrección de Pearson o de Spearman en caracteres físicos y textiles de la fibra de alpacas. *Revista de Investigaciones Veterinarias del Perú*, 33(3), 1-19. <https://doi.org/10.15381/rivep.v33i3.22908>
- Barbeau, K., Boileau, K., Sarr, F., & Smith, K. (2019). Path analysis in Mplus: A tutorial using a conceptual model of psychological and behavioral antecedents of bulimic symptoms in young adults. *The Quantitative Methods for Psychology*, 15(1), 38-53. <https://doi.org/10.20982/tqmp.15.1.p038>
- Bautista Hernández, C. M., Ganul Reich, J., López Ramírez, C., & De la Cruz Hernández, E. (2024). Un modelo de Red Bayesiana para datos cualitativos de Vaginosis Bacteriana en mujeres embarazadas. *Ideas en Ciencia e Ingeniería*, 2, 62-77. <https://doi.org/10.36677/ideaseningeneria.v2i2.24096>
- Bouza, C. N. (2023). Análisis exploratorio de datos univariados para la ciencia de datos. *Red Iberoamericana de Estudios Cuantitativos Aplicados*. <https://doi.org/10.13140/RG.2.2.26038.06722>
- Brand, Y. (2021). *Modelos de ecuaciones estructurales: conceptos y aplicaciones* [Tesis doctoral, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/bitstream/handle/unal/80064/1059705148.2021.pdf>
- Castillo, A., & Ruiz, A. (2025). Entre lo observable y lo latente: modelos de ecuaciones estructurales e investigación social. *Revista Reflexiones*, 104, 1-27. <https://doi.org/10.15517/r.v104i2.58268>
- Castro-Castrillo, C., & Duarte-Artavia, C. (2025). Bacterial vaginosis: The role of *Atopobium vaginae* and other anaerobic bacteria. *Clínica y Revisión en Ciencias Médicas*, 9, 2. <https://doi.org/10.37811/cl.rcm.v9i2.17016>
- Civelek, M. E. (2018). *Essentials of Structural Equation Modeling*. Zea E-Books, University of Nebraska–Lincoln. <https://digitalcommons.unl.edu/zeabook/64>
- Cortés, F. (2018). Observación, causalidad y explicación causal. *Perfiles Latinoamericanos*, 26(52), 1-20. <https://doi.org/10.18504/pl2652-001-2018>

- Das, S., Basanti, N., Singh, Y. A., & Kulnu, N. (2024). Pregnancy outcome in women with bacterial vaginosis. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, 13, 922-925. <https://www.ijrcog.org/index.php/ijrcog/article/view/15294>
- De la Cruz-Ruiz, F., Canul-Reich, J., Rivera-López, R., & De la Cruz-Hernández, E. (2024). Impact of data balancing a multiclass dataset before the creation of association rules to study bacterial vaginosis. *Intelligent Medicine*, 4(3), 188-199. <https://doi.org/10.1016/j.imed.2023.02.001>
- Flora, D., Crone, G., & Bell, S. (2025). Effect size interpretation in structural equation models. *Structural Equation Modeling*, 32(6), 1069-1076. <https://doi.org/10.1080/10705511.2025.2459768>
- García-Avalos, M., Canul-Reich, J., Rodríguez-Henríquez, L. M. X., & De la Cruz-Hernández, E. N. (2025). Causal Graphical Model of Bacterial Vaginosis in Pregnant Women. *Diseases*, 13, 375. <https://doi.org/10.3390/diseases13110375>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). *Manual de Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd). OmniaScience. <https://doi.org/10.3926/oms.108>
- Hernández-Gómez, H. J., Canul-Reich, J., Hernández-Ocaña, B., & De la Cruz-Hernández, E. (2023). An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis*, 27(3), 583-611. <https://doi.org/10.3233/IDA-216488>
- Jordán-Muñoz, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Revista de Investigación en Psicología Social*, 7(1), 66-71. <https://www.redalyc.org/articulo.oa?id=672371335005>
- Lepera, A. (2021). Introducción a los Modelos de Ecuaciones Estructurales y su Implementación en R mediante un ejemplo. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión Económica*, 1(1), 15-37. <https://www.economicas.uba.ar/wp-content/uploads/2016/04/Lepera-Andrea-1.pdf>
- Luzuriaga, H., Espinosa, C., Haro, A., & Ortiz, H. (2023). Histograma y distribución normal: Shapiro-Wilk y Kolmogorov Smirnov aplicado en SPSS. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4, 596-607. <https://doi.org/10.56712/latam.v4i4.1242>

- Manzano, A. (2017). Introducción a los modelos de ecuaciones estructurales. *Investigación y Educación Médica*, 7, 67-72. <https://doi.org/10.1016/j.riem.2017.11.002>
- Mendivelso, F., & Rodríguez, M. (2021). Prueba no paramétrica de correlación de Spearman. *Revista Médica Sanitas*, 24, 42-45. <https://doi.org/10.26852/01234250.578>
- Mending, W., Palmeira-de-Oliveira, A., Biber, S., & Prasauskas, V. (2019). An update on the role of *Atopobium vaginae* in bacterial vaginosis: what to consider when choosing a treatment? A mini review. *Archives of Gynecology and Obstetrics*, 300, 1-6. <https://doi.org/10.1007/s00404-019-05142-8>
- Mina-Ortiz, J. B., Franco-Macias, M. O., Santana-Mariscal, L. A., & Garcia-Ortega, M. G. (2024). Impact on maternal and fetal health of adolescent pregnant women with bacterial vaginosis. *Journal of Scientific and MQR Investigar*, 8, 5241-5264. <https://doi.org/10.56048/MQR20225.8.1.2024.5041-5264>
- Morelli-Martínez, I., & Gamboa-Miranda, S. (2022). Vaginosis bacteriana en el embarazo: últimos avances hasta la fecha. *Revista Médica Sinergia*, 7, e838. <https://doi.org/10.31434/rms.v7i7.838>
- Park, H., Shin, J., Park, J. Y., Lee, S., & Kim, Y. (2024). Prevalence of bacterial vaginosis in pregnant women in low- and middle-income countries: a systematic review and meta-analysis. *Reproductive Health Matters*, 32, 99. <https://doi.org/10.1002/rfc2.99>
- Pérez-Gómez, J. F., Canul-Reich, J., & De la Cruz-Hernández, E. (2020). Combination of Rankings as a Method for Biomarker Identification of Bacterial Vaginosis. *Research in Computing Science*, 149, 915-927. <https://dblp.org/rec/journals/racs/Perez-GomezCD20>
- Platas, V. (2021). *Contrastes de normalidad* [Tesis doctoral, Universidad de Santiago de Compostela]. <https://minerva.usc.gal/rest/api/core/bitstreams/c8efafac-12ac-446a-9443-bb3ca39fa5f0/content>
- Rashid, W., & Gupta, M. K. (2021). A Perspective of Missing Value Imputation Approaches. *Advances in Computational Intelligence and Communication Technology: Proceedings of CICT 2019*, 1086, 307-315. https://doi.org/10.1007/978-981-15-1275-9_25
- Sagaró-del-Campo, N. M., & Zamora-Matamoros, L. (2019). Métodos gráficos en la investigación biomédica de causalidad. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, 44(4), 1-9. <http://revzoilomarinellosld.cu/index.php/zmv/article/view/1846>

- Sanchez-Garcia, E. K., Contreras-Paredes, A., Martinez-Abundis, E., Garcia-Chan, D., Lizano, M., & De la Cruz-Hernandez, E. (2019). Molecular epidemiology of bacterial vaginosis and its association with genital micro-organisms in asymptomatic women. *Journal of Medical Microbiology*, 68(9), 1373-1382. <https://doi.org/10.1099/jmm.0.001044>
- Sethi, N., Narayanan, V., Saaid, R., Ahmad Adlan, A. S., Ngoi, S. T., Teh, C. S. J., Hamidi, M., & Group, W. R. (2025). Prevalence, risk factors, and adverse outcomes of bacterial vaginosis among pregnant women: a systematic review. *BMC Pregnancy and Childbirth*, 25, 40. <https://doi.org/10.1186/s12884-025-07144-8>
- Sucar, L. E. (2021). *Probabilistic Graphical Models: Principles and Applications* (2nd). Springer.
- Tortora, G. J., Funke, B. R., & Case, C. L. (2021). *Microbiology: An Introduction* (14th). Pearson. <https://www.pearson.com/en-us/subject-catalog/p/microbiology-an-introduction/P200000005614>

Capítulo 5

Conclusiones y recomendaciones generales

5.1. Conclusiones generales

En la presente investigación se analizó un conjunto de datos que incluye el diagnóstico de Vaginosis Bacteriana (VB), una alteración de la microbiota vaginal caracterizada por un desequilibrio en las comunidades bacterianas. Dicho conjunto de datos está conformado por 132 mujeres embarazadas del estado de Tabasco, quienes participaron en una campaña de embarazo saludable, entre las 4 y 24 semanas de gestación. Se identificaron bacterias relevantes asociadas con el diagnóstico de VB mediante la matriz de correlación de Pearson, entre las que destacan: *Mycoplasma hominis* (Mh), *Atopobium vaginae* (Av), *Gardnerella vaginalis* (Gv), *Megasphaera Type 1* (MT1) y *Bacterial Vaginosis Associated Bacterium Type 2* (BVAB2).

Asimismo, la aplicación del Análisis de Trayectorias (AT) permitió construir un primer modelo gráfico causal (MGC), en el que se evidencian las relaciones directas e indirectas entre las bacterias analizadas y el diagnóstico de VB. Posteriormente, mediante la matriz de correlación de Spearman, se identificaron las mismas bacterias relevantes reportadas con Pearson. A partir de ello, el segundo modelo gráfico causal permitió identificar a MT1 y BVAB2 como las bacterias con mayor influencia sobre el diagnóstico de VB, presentando efectos totales de 0.4900 y 0.5750, respectivamente. Las bacterias restantes mostraron contribuciones diferenciadas dentro del eco-

sistema microbiológico, tales como Av con un efecto total de 0.2627, Gv con 0.0482 y Mh con un efecto negativo de -0.1937.

Finalmente, se desarrollaron dos modelos gráficos causales mediante análisis de trayectorias, los cuales permitieron estudiar las transiciones entre los distintos estados de la microbiota vaginal: el primero entre la microbiota normal (VB-) y el estado intermedio (I), y el segundo entre el estado intermedio (I) y la VB positiva (VB+). Los resultados mostraron variaciones en los efectos totales de las bacterias durante cada etapa de transición. En el cambio de VB- a I, MT1 presentó un efecto positivo moderado de 0.3184, mientras que Av y Gv mostraron efectos positivos débiles de 0.0260 y 0.0388, respectivamente. En contraste, Mh presentó un efecto negativo importante de -0.4400 y BVAB2 un efecto negativo leve de -0.0188.

Durante la transición de I a VB+, Av mostró el efecto positivo más fuerte (0.6260), seguida de BVAB2 (0.3800), ambas asociadas con la progresión hacia una disbiosis avanzada. Asimismo, MT1 permaneció positivamente asociada con un efecto de 0.1164 y Mh presentó un efecto positivo de 0.1256, mientras que Gv mostró un efecto negativo leve de -0.0138.

La aplicación del análisis de trayectorias al conjunto de datos permitió comprender con mayor claridad las asociaciones entre las bacterias y el diagnóstico de VB, evidenciando que esta metodología constituye una herramienta útil para la visualización y el análisis de patrones de asociación en estudios de microbiota vaginal. Estos hallazgos contribuyen a una mejor comprensión de la dinámica microbiana asociada con la progresión de la VB en mujeres embarazadas.

5.2. Aportaciones de la investigación

Las principales aportaciones de esta investigación se resumen a continuación:

- Se aplicó el Análisis de Trayectorias (AT) como herramienta de modelado estadístico para estudiar las relaciones entre bacterias asociadas a la Vaginosis Bacteriana (VB) en mujeres embarazadas, mostrando su utilidad para analizar interacciones en estudios de microbiota vaginal.
- Se desarrollaron Modelos Gráficos Causales (MGCs) que permiten representar de manera visual las relaciones directas e indirectas entre las bacterias y el diagnóstico de VB, facili-

tando la interpretación de los resultados desde una perspectiva estadística y biológica.

- Se identificaron bacterias relevantes asociadas al diagnóstico de VB, destacando la participación de Mh, Av, Gv, MT1 y BVAB2, las cuales presentan diferentes niveles de influencia dentro de los modelos propuestos.
- Se analizaron las transiciones entre los estados de la microbiota vaginal (VB-, estado intermedio y VB+), permitiendo identificar bacterias que influyen en cada etapa del proceso de progresión de la VB.
- Desde el área de la computación, se propone un enfoque metodológico basado en el uso de técnicas estadísticas y modelos gráficos causales para el análisis de datos microbiológicos, lo que contribuye al desarrollo de herramientas analíticas para el estudio de sistemas biológicos complicados.
- Asimismo, este enfoque contribuye al desarrollo de métodos computacionales orientados al análisis de asociaciones en datos biomédicos, favoreciendo la interpretación estructurada de información complicada y brindando apoyo a la toma de decisiones en investigación.
- Los modelos propuestos constituyen una herramienta metodológica que puede contribuir al estudio de la VB y apoyar futuras investigaciones relacionadas con el análisis de microbiota vaginal.

5.3. Limitaciones del estudio

A pesar de los resultados obtenidos, este estudio presenta algunas limitaciones que deben considerarse al interpretar los modelos propuestos. En primer lugar, el análisis se realizó con un conjunto de datos correspondiente a 132 mujeres embarazadas provenientes de una región específica, lo que podría limitar la generalización de los resultados a otras poblaciones.

Asimismo, los modelos desarrollados se basan en variables observables y en el análisis de correlaciones, por lo que las relaciones identificadas deben interpretarse dentro del marco estadístico utilizado. Aunque el Análisis de Trayectorias permite estudiar efectos directos e indirectos entre

variables, los resultados dependen de la estructura teórica planteada y de las características del conjunto de datos analizado.

Finalmente, los datos utilizados en este trabajo fueron recolectados previamente por terceros y, aunque incluían diversas variables clínicas y sociodemográficas, se analizaron bajo un enfoque transversal. Esta característica metodológica limitó la posibilidad de observar la evolución de la microbiota vaginal en el tiempo y de incorporar dinámicas longitudinales que podrían enriquecer el análisis.

5.4. Recomendaciones

Como trabajo futuro, se recomienda ampliar el análisis mediante conjuntos de datos con un mayor número de muestras, con el objetivo de validar y fortalecer los modelos propuestos. De manera particular, se plantea desarrollar un estudio longitudinal que incluya el seguimiento de pacientes en diferentes etapas del embarazo, a fin de observar la evolución de la microbiota vaginal a lo largo del tiempo y profundizar en el análisis de sus dinámicas microbianas. Asimismo, resulta de interés explorar otros enfoques de modelado de ecuaciones estructurales que complementen el estudio de las interacciones bacterianas.

Finalmente, los modelos gráficos causales desarrollados en esta investigación representan una herramienta de apoyo para investigadores y especialistas en el estudio de la microbiota vaginal, aportando elementos para una mejor comprensión de los procesos asociados al desarrollo y progresión de la Vaginosis Bacteriana.

5.5. Publicaciones derivadas de la investigación

Como resultado del desarrollo de la presente investigación, se generaron artículos que están publicados en revistas especializadas. Cabe señalar que los artículos integrados en esta tesis corresponden a versiones sometidas a evaluación editorial. Las versiones finales publicadas incorporan modificaciones derivadas del proceso de arbitraje por pares.

A continuación se presentan las referencias correspondientes:

- García-Avalos, M., Canul-Reich, J., Rodríguez-Henríquez, L.M., De la Cruz-Hernández, E.

Título del artículo: Hacia un Modelo Gráfico Causal para el Análisis de la Vaginosis Bacteriana. Revista: Ideas en Ciencias de la Ingeniería, Vol. 3, Núm. 1, 2025, pp. 22–34. DOI: <https://doi.org/10.36677/rici.v3i1.25016>

- García-Avalos, M., Canul-Reich, J., Rodríguez-Henríquez, L.M.X., De la Cruz-Hernández, E.N. Title: Causal Graphical Model of Bacterial Vaginosis in Pregnant Women. Journal: Diseases, Vol. 13, 2025, Article 375. DOI: <https://doi.org/10.3390/diseases13110375>
- García-Avalos, M., Canul-Reich, J., Rodríguez-Henríquez, L.M.X., De la Cruz-Hernández, E.N. Title: Causal Graphical Models for Transition from Healthy Vaginal Microbiota to Bacterial Vaginosis in Pregnant Women. Journal: BioMedInformatics, Vol. 6, 2026, Article 32. DOI: <https://doi.org/10.3390/biomedinformatics6030032>

Universidad Juárez Autónoma de Tabasco.
México.

Alojamiento de la Tesis en el Repositorio Institucional	
Título de la tesis:	Modelo gráfico causal para representar las bacterias coexistentes detonantes de vaginosis bacteriana en mujeres embarazadas
Autor:	Maricela García Avalos
ORCID:	https://orcid.org/0009-0007-0643-8977
Resumen:	<p>La vaginosis bacteriana (VB) es una alteración de la microbiota vaginal caracterizada por un desequilibrio en las comunidades bacterianas, cuya composición puede variar entre mujeres y evolucionar a través de distintos estados. Comprender las interacciones entre las bacterias asociadas y su influencia en el desarrollo de la VB es fundamental para mejorar su interpretación clínica y biológica. En esta investigación se analizan dichas relaciones mediante la aplicación del Análisis de Trayectorias (AT), a partir del cual se construyen Modelos Gráficos Causales (MGC), que permiten representar y estudiar las relaciones entre especies bacterianas involucradas en esta infección. El análisis se realizó con datos de 132 mujeres embarazadas entre 4 y 24 semanas de gestación. Inicialmente se construyó un modelo basado en la matriz de correlación de Pearson para identificar bacterias asociadas al diagnóstico de VB. Posteriormente, se desarrollaron MGC basados en la correlación de Spearman para examinar las relaciones causales entre bacterias y analizar las transiciones entre microbiota normal (VB-), estado intermedio (I) y VB positiva (VB+). Los resultados destacan la participación de <i>Mycoplasma hominis</i> (Mh), <i>Atopobium vaginae</i> (Av), <i>Gardnerella vaginalis</i> (Gv), <i>Megasphaera</i> Tipo 1 (MT1) y BVAB2. Además, se observó que MT1 contribuye a la transición desde VB- hacia el estado I, mientras que Av y BVAB2 influyen en la transición desde I hacia VB+. Estos modelos contribuyen a comprender mejor la dinámica bacteriana asociada a la VB.</p>
Palabras clave:	Análisis de Trayectorias, Matriz de Correlación, Métricas Estadísticas, Bacterias Anaerobias
Referencias citadas:	En la siguiente página se muestran las referencias.

Bibliografía

- Abundis, E. M., Hernandez-Landero, F., Escobar-Calderon, G., Gomez-Crisostomo, N., Contreras-Paredes, A., & de la Cruz-Hernandez, E. (2024). Gene expression of cardiovascular risk markers in mononuclear cells of pregnant women in relation to plasma leptin and homocysteine levels: A cross-sectional study. *International Journal of Gynaecology and Obstetrics*, 165, 350-360. <https://doi.org/10.1002/ijgo.15302>
- Álvarez-Vega, M., Quirós-Mora, L. M., & Cortés-Badilla, M. V. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, 5(8), 1-12. <https://doi.org/10.31434/rms.v5i8.557>
- Alwateer, M., Atlam, E., Abd, M., Ghoneim, O., & Gad, I. (2024). Missing Data Imputation: A Comprehensive Review. *Journal of Computer and Communications*, 12, 53-75. <https://doi.org/10.4236/jcc.2024.1211004>
- Apaza, E., Cazorla, S., Condori, C., Arpasi, F., Tumi, I., Yana, W., & Quispe, J. (2022). La Correlación de Pearson o de Spearman en caracteres físicos y textiles de la fibra de alpacas. *Revista de Investigaciones Veterinarias del Perú*, 33(3), 1-19. <https://doi.org/10.15381/rivep.v33i3.22908>
- Ayala, J. (2023). Transformaciones. <https://rpubs.com/JairoAyala/TNM>
- Barbeau, K., Boileau, K., Sarr, F., & Smith, K. (2019). Path analysis in Mplus: A tutorial using a conceptual model of psychological and behavioral antecedents of bulimic symptoms in young adults. *The Quantitative Methods for Psychology*, 15(1), 38-53. <https://doi.org/10.20982/tqmp.15.1.p038>
- Bautista Hernández, C. M., Canul Reich, J., López Ramírez, C., & De la Cruz Hernández, E. (2024). Un modelo de Red Bayesiana para datos cualitativos de Vaginosis Bacteriana en mujeres embarazadas. *Ideas en Ciencia e Ingeniería*, 2, 62-77. <https://doi.org/10.36677/ideaseningeneria.v2i2.24096>
- Benavent-Nuñez, D., Colomer-Mascaró, J., Quecedo-Gutiérrez, L., Gol-Montserrat, J., & Del Llano-Señaris, J. E. (2020). *Inteligencia Artificial y decisiones Clínicas*. Fundación Gaspar Casal.
- Bouza, C. N. (2023). Análisis exploratorio de datos univariados para la ciencia de datos. *Red Iberoamericana de Estudios Cuantitativos Aplicados*. <https://doi.org/10.13140/RG.2.2.26038.06722>

- Brand, Y. (2021). *Modelos de ecuaciones estructurales: conceptos y aplicaciones* [Tesis doctoral, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/bitstream/handle/unal/80064/1059705148.2021.pdf>
- Cáceres, M., & Ponce, R. (2023). Dependencia emocional y su relación con el riesgo suicida en adultos jóvenes. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4. <https://doi.org/10.56712/latam.v4i1.247>
- Cameron, C., Dion, M., Broce, J., Ojo, D., Drobina, E., Louis, A., Girbert, J., Fang, R., & Parker, I. (2023). Ethnic disparity in diagnosing asymptomatic bacterial vaginosis using machine learning. *npj Digital Medicine*, 6, 1-10. <https://doi.org/10.1038/s41746-023-00953-1>
- Castillo, A., & Ruiz, A. (2025). Entre lo observable y lo latente: modelos de ecuaciones estructurales e investigación social. *Revista Reflexiones*, 104, 1-27. <https://doi.org/10.15517/rr.v104i2.58268>
- Castro-Castrillo, C., & Duarte-Artavia, C. (2025). Bacterial vaginosis: The role of *Atopobium vaginae* and other anaerobic bacteria. *Clínica y Revisión en Ciencias Médicas*, 9, 2. https://doi.org/10.37811/cl_rcm.v9i2.17016
- Civelek, M. E. (2018). *Essentials of Structural Equation Modeling*. Zea E-Books, University of Nebraska–Lincoln. <https://digitalcommons.unl.edu/zeabook/64>
- Contreras-Cueva, A. B., Luna-Plascencia, N. R., Macías-Álvarez, P., & González-Robles, N. M. (2024). Imputación de datos faltantes en encuestas en hogares. *Memorias de la Décima Cuarta Conferencia Iberoamericana de Complejidad, Informática y Cibernética*, 310-316. <https://doi.org/10.54808/CICIC2024.01.310>
- Cortés, F. (2018). Observación, causalidad y explicación causal. *Perfiles Latinoamericanos*, 26(52), 1-20. <https://doi.org/10.18504/pl2652-001-2018>
- Das, S., Basanti, N., Singh, Y. A., & Kulnu, N. (2024). Pregnancy outcome in women with bacterial vaginosis. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, 13, 922-925. <https://www.ijrcog.org/index.php/ijrcog/article/view/15294>
- De la Cruz-Ruiz, F., Canul-Reich, J., Rivera-López, R., & De la Cruz-Hernández, E. (2024). Impact of data balancing a multiclass dataset before the creation of association rules to study bacterial vaginosis. *Intelligent Medicine*, 4(3), 188-199. <https://doi.org/10.1016/j.imed.2023.02.001>
- Flora, D., Crone, G., & Bell, S. (2025). Effect size interpretation in structural equation models. *Structural Equation Modeling*, 32(6), 1069-1076. <https://doi.org/10.1080/10705511.2025.2459768>
- García, M., Canul, J., Rodríguez, L., & Cruz, E. (2025). Hacia un Modelo Gráfico Causal para el Análisis de la Vaginosis Bacteriana. *Ideas en Ciencia e Ingeniería*, 3, 22-34. <https://doi.org/10.36677/rici.v3i1.25016>
- García-Avalos, M., Canul-Reich, J., Rodríguez-Henríquez, L. M. X., & De la Cruz-Hernández, E. N. (2025). Causal Graphical Model of Bacterial Vaginosis in Pregnant Women. *Diseases*, 13, 375. <https://doi.org/10.3390/diseases13110375>

- González, K., Guerra, F., Villeda, G., & López, M. (2024). Frecuencia y detección molecular de *Gardnerella vaginalis* en una institución de tercer nivel. *Revista Perinatología y Reproducción Humana*, 37, 91-98. <https://doi.org/10.24875/PER.23000020>
- González, N., Santisteban, A., Ortiz, Y., Pérez, D., & González, M. (2019). Factores de riesgo asociados a infección vaginal en mujeres embarazadas. *Multimed. Revista Médica Granma*, 23(3), 430-446.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). *Manual de Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd). OmniaScience. <https://doi.org/10.3926/oms.108>
- Hamasha, M., Ali, H., Hamasha, S., & Ahmed, A. (2022). Ultra-fine transformation of data for normality. *Heliyon*, 8, e09370. <https://doi.org/10.1016/j.heliyon.2022.e09370>
- Harrison, C. J., & Sidey-Gibbons, C. J. (2021). Machine learning in medicine: a practical introduction to natural language processing. *BMC Medical Research Methodology*, 21(1), 158. <https://doi.org/10.1186/s12874-021-01347-1>
- Hernández-Gómez, H. J., Canul-Reich, J., Hernández-Ocaña, B., & De la Cruz-Hernández, E. (2023). An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis*, 27(3), 583-611. <https://doi.org/10.3233/IDA-216488>
- Iglesias Labraca, J. M., & Reche Lorite, F. (2021). Modelos de Ecuaciones Estructurales. *Universidad de Almería*, 1-66. <https://repositorio.ual.es/bitstream/handle/10835/13177/IGLESIAS%20LABRACA,%20JUAN%20MIGUEL.pdf?sequence=1>
- Jahuey-Martinez, F. J., Herrera-Ojeda, J. B., & Paredes-Sánchez, F. A. (2022). El programa R: una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria*, 23(4). <https://doi.org/10.22201/cuaieed.16076079e.2022.23.4.4>
- Jordán-Muiños, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Revista de Investigación en Psicología Social*, 7(1), 66-71. <https://www.redalyc.org/articulo.oa?id=672371335005>
- Lepera, A. (2021). Introducción a los Modelos de Ecuaciones Estructurales y su Implementación en R mediante un ejemplo. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión Económica*, 1(1), 15-37. <https://www.economicas.uba.ar/wp-content/uploads/2016/04/Lepera-Andrea-1.pdf>
- Luzuriaga, H., Espinosa, C., Haro, A., & Ortiz, H. (2023). Histograma y distribución normal: Shapiro-Wilk y Kolmogorov Smirnov aplicado en SPSS. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4, 596-607. <https://doi.org/10.56712/latam.v4i4.1242>
- Madrid, A. E., Valenzuela-Ruiz, S. M., Batanero, C., & Garzón-Guerrero, J. A. (2022). Interpretación del diagrama de caja por estudiantes universitarios de ciencias de la actividad física y deporte. *Educación Matemática*, 34(3). <https://doi.org/10.24844/em3403.10>
- Manzano, A. (2017). Introducción a los modelos de ecuaciones estructurales. *Investigación y Educación Médica*, 7, 67-72. <https://doi.org/10.1016/j.riem.2017.11.002>

- Márquez-Díaz, J. (2020). Inteligencia artificial y Big Data como soluciones frente a la COVID-19. *Revista de Bioética y Derecho*, 50, 315-331. <https://doi.org/10.1344/rbd2020.50.31643>
- Mejía-Peñafiel, E. F., Aguilar-Reyes, J. E., Bolaños-Logroño, P. F., & López-Ortega, J. R. (2024). Métodos de imputación por regresión, imputación por moda, imputación múltiple y árboles de decisión para variables categóricas en perspectiva comparada. *Revista De La Universidad Del Zulia*, 15(43), 541-555. <https://doi.org/10.46925/rdluz.43.29>
- Mendivelso, F., & Rodríguez, M. (2021). Prueba no paramétrica de correlación de Spearman. *Revista Médica Sanitas*, 24, 42-45. <https://doi.org/10.26852/01234250.578>
- Mending, W., Palmeira-de-Oliveira, A., Biber, S., & Prasauskas, V. (2019). An update on the role of *Atopobium vaginae* in bacterial vaginosis: what to consider when choosing a treatment? A mini review. *Archives of Gynecology and Obstetrics*, 300, 1-6. <https://doi.org/10.1007/s00404-019-05142-8>
- Mendoza-Sánchez, D., Quintana-Zúñiga, Á. A., & Díaz-Quiñonez, A. (2023). La importancia de la investigación científica en medicina. *Revista Médica Sinergia, Atención Familiar UNAM*, 24(3), 224-227. <https://doi.org/10.22201/fm.14058871p.2023.3>
- Mina-Ortiz, J. B., Franco-Macias, M. O., Santana-Mariscal, L. A., & Garcia-Ortega, M. G. (2024). Impact on maternal and fetal health of adolescent pregnant women with bacterial vaginosis. *Journal of Scientific and MQR Investigar*, 8, 5241-5264. <https://doi.org/10.56048/MQR20225.8.1.2024.5041-5264>
- Molina-Arias, M., Ochoa-Sangrador, C., & Ortega-Páez, E. (2021). Correlación, Modelos de regresión. *Evidencia Pediátrica*, 17(2), 17-25. <https://evidenciasenpediatria.es/articulo/7827/correlacion-modelos-de-regresion>
- Morelli-Martínez, I., & Gamboa-Miranda, S. (2022). Vaginosis bacteriana en el embarazo: últimos avances hasta la fecha. *Revista Médica Sinergia*, 7, e838. <https://doi.org/10.31434/rms.v7i7.838>
- Park, H., Shin, J., Park, J. Y., Lee, S., & Kim, Y. (2024). Prevalence of bacterial vaginosis in pregnant women in low- and middle-income countries: a systematic review and meta-analysis. *Reproductive Health Matters*, 32, 99. <https://doi.org/10.1002/rfc2.99>
- Pérez-Gómez, J. F., Canul-Reich, J., Hernández-Torruco, J., & Hernández-Ocaña, B. (2020). Predictor selection for bacterial vaginosis diagnosis using decision tree and relief algorithms. *Applied Sciences*, 10(9), 3291. <https://doi.org/10.3390/app10093291>
- Pérez-Gómez, J. F., Canul-Reich, J., & De la Cruz-Hernández, E. (2020). Combination of Rankings as a Method for Biomarker Identification of Bacterial Vaginosis. *Research in Computing Science*, 149, 915-927. <https://dblp.org/rec/journals/rcs/Perez-GomezCD20>
- Platas, V. (2021). *Contrastes de normalidad* [Tesis doctoral, Universidad de Santiago de Compostela]. <https://minerva.usc.gal/rest/api/core/bitstreams/c8efafac-12ac-446a-9443-bb3ca39fa5f0/content>

- Poaquiza, L., & Gavilanes, G. (2022). El estrés y su relación con la inteligencia emocional en docentes universitarios. *Ciencia Latina Revista Científica Multidisciplinaria*, 6, 1357-1372. <https://doi.org/10.37811/cl.rcm.v6i3.2301>
- Rashid, W., & Gupta, M. K. (2021). A Perspective of Missing Value Imputation Approaches. *Advances in Computational Intelligence and Communication Technology: Proceedings of CICT 2019*, 1086, 307-315. https://doi.org/10.1007/978-981-15-1275-9_25
- Romero-Herrero, D., & Andreu-Domingo, A. (2016). Vaginosis Bacteriana. *Enfermedades Infecciosas y Microbiología Clínica*, 34(3), 14-18.
- Sagaró-del-Campo, N. M., & Zamora-Matamoros, L. (2019). Métodos gráficos en la investigación biomédica de causalidad. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, 44(4), 1-9. <http://revzoilomarinello.sld.cu/index.php/zmv/article/view/1846>
- Salas, J., Angulo, L., & Mendez, E. (2022). Vaginosis bacteriana: actualización y novedad terapéutica. *UCIMED*, 5(6), 85-93. <https://doi.org/10.34192/cienciaysalud.v5i6.387>
- Sanchez-Garcia, E. K., Contreras-Paredes, A., Martinez-Abundis, E., Garcia-Chan, D., Lizano, M., & De la Cruz-Hernandez, E. (2019). Molecular epidemiology of bacterial vaginosis and its association with genital micro-organisms in asymptomatic women. *Journal of Medical Microbiology*, 68(9), 1373-1382. <https://doi.org/10.1099/jmm.0.001044>
- Sethi, N., Narayanan, V., Saaid, R., Ahmad Adlan, A. S., Ngoi, S. T., Teh, C. S. J., Hamidi, M., & Group, W. R. (2025). Prevalence, risk factors, and adverse outcomes of bacterial vaginosis among pregnant women: a systematic review. *BMC Pregnancy and Childbirth*, 25, 40. <https://doi.org/10.1186/s12884-025-07144-8>
- Sucar, L. E. (2021). *Probabilistic Graphical Models: Principles and Applications* (2nd). Springer.
- Tortora, G. J., Funke, B. R., & Case, C. L. (2021). *Microbiology: An Introduction* (14th). Pearson. <https://www.pearson.com/en-us/subject-catalog/p/microbiology-an-introduction/P200000005614>