

Un modelo de Red Bayesiana para datos cualitativos de Vaginosis Bacteriana en mujeres embarazadas

A Red Bayesian Model for Qualitative Data of Bacterial Vaginosis in Pregnant

Bautista Hernández, C. M.^{1*} , Canul Reich, J.¹ , López Ramírez, C.¹, De la Cruz Hernández, E.¹
(*carlos.bta@icloud.com)

Recibido: 13 de junio de 2024 Aceptado: 28 de junio de 2024

RESUMEN

La Vaginosis Bacteriana BV es un factor de riesgo para el parto prematuro. La Organización Mundial de la Salud OMS, destaca que a pesar de que la BV ya se considera una afección más común y de no tratarse puede provocar problemas en el embarazo como aborto espontáneo y parto prematuro así como mayor riesgo de otras infecciones de transmisión sexual. Las técnicas computacionales que parten de la Inteligencia Artificial como Redes Bayesianas Causales CBNs, se han utilizado en diferentes estudios con el objetivo de análisis de diagnóstico de enfermedades. Este estudio tiene la finalidad de compartir cómo una CBN puede identificar que bacterias coexisten y la relación que tienen frente a un diagnóstico

positivo DxBV+. Los datos que se utilizaron para este estudio fueron una recolección de datos reales por parte de personal investigador de la División Académica Multidisciplinaria de Comalcalco DAMC formando un *dataset* de casos positivos, negativos e indeterminados de BV en mujeres embarazadas. El resultado que se obtuvo dio un modelo gráfico causal que es presentado como una CBN y su relación de probabilidad condicional que detalla el porcentaje de relación con el DxBV+.

Palabras clave: *Vaginosis Bacteriana BV, Mujeres Embarazadas, Infección Genital, Redes Bayesianas BN.*

¹Universidad Juárez Autónoma de Tabasco. Carr. Cunduacan a Jalpa de Méndez, KM 1, C.P. 86690.



ABSTRACT

BV Bacterial Vaginosis is a risk factor for preterm birth. The World Health Organization WHO, highlights that although BV is already considered a more common condition and if left untreated it can cause problems in pregnancy such as spontaneous abortion and premature birth as well as an increased risk of other sexually transmitted infections. Artificial Intelligence computational techniques such as Causal Bayesian Networks CBNs have been used in different studies with the objective of disease diagnostic analysis. This study aims to share how a CBN can identify which bacteria coexist and the relationship they have with a positive DxBV+ diagnostic. The data that was used for this study is based on a collection of real data by research personnel from the Multidisciplinary Academic Division of Comalcalco DAMC, forming a dataset of positive, negative and indeterminate cases of BV in pregnant women. The result obtained gave a causal graphic model that is presented as a CBN and its conditional probability table that details the percentage of relationship with DxBV+.

Keywords: *BV Bacterial Vaginosis, Genital Infection, Pregnant Women, Bayesian networks.*

INTRODUCCIÓN

La Vaginosis Bacteriana (BV, por sus siglas en inglés) es una infección genital femenina de las más comunes y el impacto sobre la salud de una mujer es significativo. La BV es una enfermedad genital que afecta tanto a las mujeres sanas y embarazadas [1]. Se sabe que hay diferentes condiciones genéticas y otras condiciones de ambiente en las que puede iniciar la infección. Un elevado pH vaginal facilita el crecimiento de organismos de transmisión sexual [2].

Se tienen indicios que enfermedades como la clamidia, el Virus del Papiloma Humano VPH, el Síndrome de Inmunodeficiencia Humana VIH y otras de transmisión sexual pueden estar en ocasiones asociadas a la BV [3]. La definición en la que coinciden en diversos artículos es que la infección que comienza a dar lugar a la BV se origina debido a un disturbio en la flora vaginal normal y la interacción de diversos agentes patógenos [4].

La BV se conoce que es la causa más común de flujo vaginal en mujeres en edad reproductiva y está asociada a una gran cantidad de condiciones incluyendo infección postquirúrgica, enfermedad pélvica inflamatoria, parto pretérmino y enfermedades como el VPH y VIH [5]. Debido a los problemas que ocasiona la BV se plantea experimentar cómo una Red Bayesiana Causal puede dar resultados sobre qué bacterias coexisten en el diagnóstico positivo de la BV.

La Red Bayesiana Causal es un algoritmo usado en *Machine Learning* (ML) con la cual pueden crearse modelos gráficos. El uso de los modelos gráficos ayuda a realizar modelos predictivos [6] mediante un

gran procesamiento de datos [7].

Los datos que se usan en este experimento fueron recolectados de pacientes por parte de personal de investigación de la División Académica Multidisciplinaria de Comalcalco (DAMC) de la Universidad Juárez Autónoma de Tabasco UJAT [8]. El *dataset* contiene tres clases: positivos, negativos e indeterminados para el diagnóstico positivo de la BV en mujeres embarazadas.

Se realizaron los experimentos en búsqueda de causas detonantes que originan la Vaginosi s Bacteriana Positiva BV+ en el *dataset* mencionado anteriormente. Como resultado se obtuvo un Modelo Gráfico Causal que fue creado con la intervención del algoritmo *Expectation – Maximization* (EM) para aprender la estructura de la Red Bayesiana y el algoritmo para Redes Bayesianas Causales conocido como *Fast Causal Inference* (FCI) para optimizar las probabilidades.

MATERIALES Y MÉTODOS

Redes Bayesianas

La creación de las Redes Bayesianas (BN, por sus siglas en inglés) tienen como fundamento el teorema de bayes para identificar el grado de relaciones entre las clases. El Teorema de Bayes tiene sus principios en los años 1700 a través de la transformación de diversas teorías de probabilidad [9]. Fue así que Tomas Bayes (1701-1761) dio origen a este teorema basándose en las investigaciones anteriores. Las BN se representan con un Gráfico Acíclico Dirigido (DAG, por sus siglas en inglés).

Teorema de Bayes

El Teorema de Bayes tiene sus principios en el Teorema de la Probabilidad Total: la probabilidad que ocurran los sucesos A y B , es la probabilidad de que ocurra A multiplicada por la probabilidad de que ocurra B , que está condicionada por la ocurrencia de A , esto es la fórmula de la Ec. (1):

$$P(A \cap B) = P(A) \cdot P(B|A) \quad (1)$$

La fórmula mostrada en la Ec. (1) es la base para la fórmula mostrada en la Ec. (2) [9]:

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)} \quad (2)$$

Donde:

$P(A|B)$ es la probabilidad de que ocurra un evento A dado que ha ocurrido un evento B , también conocida como la probabilidad posterior.

$P(A)$ es la probabilidad de que ocurra un evento A sin tener en cuenta ninguna información sobre B conocida también como probabilidad previa o a priori.

$P(B|A)$ es la probabilidad de que ocurra el evento B , dado que ha ocurrido el evento A , conocida como probabilidad condicional.

$P(B)$ es la probabilidad de que ocurra el evento B independientemente de cualquier otra información. También conocida como probabilidad marginal [9].

Interpretación y aplicaciones del Teorema de Bayes

El Teorema de Bayes describe la relación entre la probabilidad condicional de dos eventos y permite actualizar la probabilidad de una hipótesis basándose en nueva evidencia. La fórmula de Bayes $P(A|B)$ permite calcular la probabilidad de A dado que B ha ocurrido, o de manera inversa la probabilidad de B dado que ha ocurrido el evento A , $P(B|A)$. El Teorema de Bayes se utiliza ampliamente en medicina y otras disciplinas [9].

Inferencia Causal

La Inferencia Causal (CI, por sus siglas en inglés) puede parecer simple, pero es una tarea científica compleja ya que se basa en la triangulación de evidencias de múltiples fuentes y en la aplicación de una variedad de enfoques metodológicos. La CI ha sido atribuida a cómo poder predecir resultados mediante eventos causales. Judea Pearl menciona que a través de la estadística se han proporcionado herramientas de predicción de probabilidades [10].

El objetivo de la CI es inferir no solo creencias o probabilidades en condiciones estáticas, sino también la dinámica de las creencias en condiciones cambiantes, cambios inducidos por tratamientos, nuevas políticas u otras intervenciones externas [11]. La CI recaba conocimiento que toma para la predicción de sucesos mediante el almacenamiento y procesamiento en una memoria.

Modelos Gráficos Causales

El estudio obtuvo como resultado la creación de un Modelo Gráfico Causal (MGC), mediante la combinación de BN y CI. Las dos técnicas aplicadas dieron un resultado de Redes Bayesianas Causales (CBN, por sus siglas en inglés) [3].

Las CBN se representan mediante un MGC el cual es una representación visual y matemática de las relaciones causales que se pueden encontrar entre un conjunto de variables. La creación de un MGC requiere grafos con inferencia causal, además cuenta con nodos que representan variables por ejemplo A, B, C y aristas que representan las relaciones causales [11], ver Ec. (3):

$$A \rightarrow B \rightarrow C \quad (3)$$

Donde: A causa B y B causa C

Redes Bayesianas Causales

Las CBNs se pueden interpretar con DAGs, donde los nodos representan variables y los arcos relaciones causales. En un MGC las CBNs representan relaciones causas-efecto, por lo que implican supuestos más fuertes que tener simplemente una BN.

La estructura de una CBN se puede construir de dos maneras. La primera manera, es aplicando algoritmos para las BNs y realizar el DAG de manera manual con base en estas relaciones; la segunda, es aplicando un algoritmo que realice las relaciones entre variables con base en los cálculos de probabilidad. A las clases objetivo se les debe fijar el valor de una o más variables y establecer su efecto en las demás, esto es el razonamiento causal [3].

Algoritmo Expectation-Maximization (EM)

EM es una técnica iterativa para estimar los parámetros de los modelos probabilísticos que permite generar una estructura de BN aplicándolo en herramientas de desarrollo de software [12]. El algoritmo EM se utiliza para crear un modelo que relaciona los datos observados y ausentes con elecciones de decisores. El modelo predice las elecciones producidas de los datos observados [13].

La idea básica del algoritmo EM es formar la estructura con los mejores parámetros de un modelo, cuyos pasos en este estudio son:

- **Inicialización:** Se inicia con suposiciones aleatorias o neutrales sobre las probabilidades.
- **Expectation (E):** Se utilizan las probabilidades condicionales de las variables con las que se cuenta para calcular las probabilidades marginales de las variables que parecen no tener una intervención. Por ejemplo, si se conoce la probabilidad de que llueva dado el clima y la temperatura, se procede a calcular la probabilidad de que llueva sin tener en cuenta estos factores.
- **Maximization (M):** Se utilizan las estimaciones de las probabilidades marginales para actualizar las probabilidades condicionales de las variables. Por ejemplo, si se observa que la lluvia es más probable cuando hace frío y está nublado, se ajustan las probabilidades condicionales de lluvia dada la temperatura y el clima para reflejar esto.
- **Repetir:** Se repiten los pasos E y M hasta que las probabilidades condicionales converjan a valores estables. Para mayor detalle ver [11].

Dataset

En el *dataset* se utiliza el algoritmo EM para generar la estructura con base en las probabilidades condicionales de todas las variables cualitativas. El conjunto de datos que se utilizó para este estudio proviene de la recolección de datos del investigador científico Dr. Erick de la Cruz Hernández perteneciente DAMC de la UJAT.

El *dataset* se encuentra constituido de 11,484 instancias con una distribución de 87 columnas y 132 filas. Se encuentran dos categorías diferentes: variables cualitativas que son 59 incluyendo dos variables clases u objetivo y variables cuantitativas que son 28. La variable clase se encuentra con el nombre de DxVB+ que es el diagnóstico positivo para la Vaginosis Bacteriana y se encuentra clasificada de la siguiente manera: para los casos positivos se representan con 1, para casos negativos con 2 y, finalmente, para casos indeterminados con 3.

Software

Se utilizó el lenguaje de programación R, el cual se ejecutó en el Entorno de Desarrollo Integrado (IDE, por sus siglas en inglés), *RStudio* en su versión 2023.12.0.369.

ESTUDIO EXPERIMENTAL

El método utilizado en la experimentación consta de las etapas a continuación descritas:

- Exploración de datos

En la variable clase se encuentran tres categorías descritas en la Tabla 1.

Tabla 1. Análisis de las instancias de la variable clase.

Casos	DxVB+
Positivos (1)	32 registros
Negativos (2)	5 registros
Indeterminados (3)	95 registros

- Preprocesamiento de los datos

En esta fase se realizaron los procedimientos para llevar a cabo la adecuación de las variables del *dataset*. Se utilizaron cuatro procedimientos principales que constan de: a) selección de datos, b) limpieza de los datos, c) transformación de los datos, d) imputación e integración de datos faltantes.

- a) Selección de los datos. El *dataset* que constaba de 87 variables y se redujo a 77 variables, quedo conformado por 28 variables cuantitativas y 47 variables cualitativas que fueron las utilizadas. Se descartaron variables como datos generales de los pacientes y otras clases de estudio como el VPH.
- b) Limpieza de datos: Se trataron datos con información anómala o con errores de captura.
- c) Transformación de datos: Algunas variables numéricas necesitaron ser transformadas a variables categóricas para poder trabajar con los datos cualitativos presentes en el *dataset*.
- d) Imputación e integración de los datos: Se tomaron las variables cualitativas y se realizó un procedimiento con base en la técnica estadística MODA [12], para rellenar los datos faltantes.

La MODA aplicada, se realizó con relación a la variable clase, es decir a la variable del diagnóstico positivo de la BV. En la Tabla 2 se muestra la lógica que se precisó en este método, en dos variables cualitativas con un ejemplo de 11 instancias por variable.

Tabla 2. Ejemplo de imputación, datos originales de *dataset* de BV.

Comezón genital	Ardor	Diagnóstico BV+
2	2	3
NA	NA	1
NA	NA	3
NA	NA	3
1	2	3
NA	NA	3
2	2	3
2	2	3
2	2	1
2	2	2
2	2	2

Con base en la Tabla 2, cuando el diagnóstico BV+ es 1, 2 o 3, la MODA para Comezón genital es 2 y para Ardor es 2, por lo que se procede a la imputación mostrados en la Tabla 3.

Tabla 3. Ejemplo de imputación, datos procesados.

Comezón genital	Ardor	Diagnóstico BV+
2	2	3
2	2	1
2	2	3
2	2	3
1	2	3
2	2	3
2	2	3
2	2	3
2	2	1
2	2	2
2	2	2

En la Tabla 3 se integran los datos al *dataset* y se concluye con el ejemplo de lo realizado en la etapa de preprocesamiento o preparación de los datos.

■ Experimentación con los datos

En la experimentación se tomó el resultado final del preprocesamiento de los datos descritos en la sección anterior. Se implementó el algoritmo EM con las 47 variables cualitativas del *dataset* para obtener la BN, el algoritmo EM se utilizó con los parámetros que tiene predefinidos por defecto. Para el análisis de resultados se ajustaron los parámetros para obtener como resultado solo las variables que se encuentran con un rango mayor al 50 % de relación de la probabilidad condicional que calcula el algoritmo EM.

El algoritmo EM hace los cálculos de probabilidad condicional entre las variables cualitativas del *dataset* y la variable clase identificada como diagnóstico positivo de la BV. Inicialmente se obtuvo una BN creada con la implementación del algoritmo EM en combinación con todas las variables cualitativas del *dataset*. El resultado inicial de esta BN experimentada se muestra en la Figura 1.

En la Figura 1 se pueden apreciar nodos y arcos dirigidos. La dirección del arco sugiere una interpretación causal o probabilística entre los nodos. Considerando a la variable *Lactobacillus inners* como el nodo (A) y a la variable Dolor como el nodo (B), se interpreta que el nodo A tiene un efecto causal o una influencia directa sobre el nodo B.

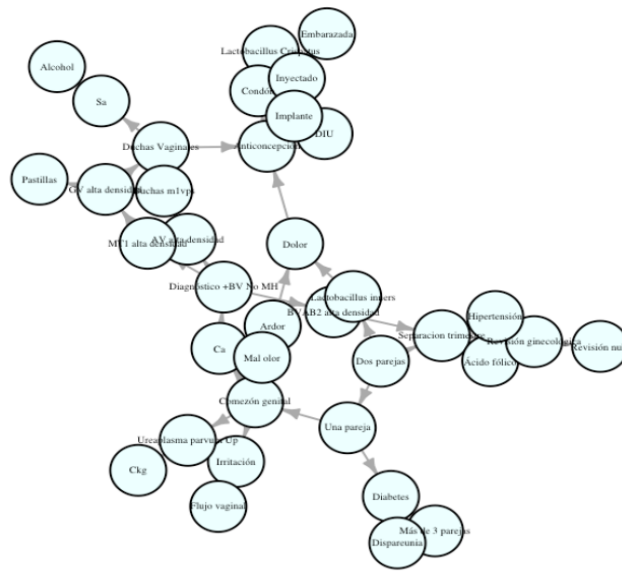


FIGURA 1. Gráfico de Red Bayesiana implementando el algoritmo EM que relaciona todas las variables cualitativas presentadas en el dataset.

En la Figura 2 se visualizan las relaciones de probabilidad creadas por el algoritmo EM en las variables cualitativas del *dataset*, representadas en cuatro columnas. La columna *from* es la primera variable relacionada, la columna *to* muestra con qué variable se está haciendo la relación desde *from* y finalmente las columnas *strength* y *direction* muestran los puntajes de relación de acuerdo a la probabilidad condicional y resultados de los cálculos dados por el algoritmo EM. La Figura 2 son los resultados de probabilidad que tienen las variables expresadas en el gráfico de la Figura 1.

En la Figura 2 se encuentra el porcentaje de relación entre las variables de la columna *from* y *to*. El porcentaje viene representado en las columnas *strength* y *direction*; para la variable AV alta densidad se puede apreciar que tiene 0.985 en los cálculos de probabilidad con respecto a la variable de Diagnóstico +BV, esto indica que se tiene una relación de 98.5 % con respecto a DxBV+.

##	from	to	strength	direction
## 40	Diagnóstico +BV No MH	AV alta densidad	0.985	0.6802030
## 42	Diagnóstico +BV No MH	MT1 alta densidad	0.960	0.7838542
## 43	Diagnóstico +BV No MH	BVAB2 alta densidad	0.775	0.6709677
## 46	Separacion trimestre	Hipertensión	0.615	0.7601626
## 49	Separacion trimestre	Revisión ginecológica	0.545	0.7064220
## 55	Separacion trimestre	Ácido fólico	0.570	0.8201754
## 96	Diabetes	Más de 3 parejas	0.650	0.7769231
## 116	Diabetes	Dispareunia	0.695	0.5647482
## 265	Revisión ginecológica	Revisión nula	1.000	0.8800000
## 353	Una pareja	Dos parejas	1.000	0.8525000
## 426	Dos parejas	Lactobacillus inners	0.520	0.9807692
## 661	Anticoncepción	Condón	1.000	0.8425000
## 663	Anticoncepción	Inyectado	0.995	0.7939698
## 812	Inyectado	Lactobacillus Crispatus	0.595	0.7394958
## 969	Duchas Vaginales	Duchas mlvps	1.000	0.9000000
## 1057	Comezón genital	Ardor	0.715	0.7307692
## 1058	Comezón genital	Irritación	0.650	0.7807692
## 1065	Comezón genital	Ureaplasma parvum Up	0.680	0.7720588
## 1066	Comezón genital	Ca	0.540	0.7129630
## 1102	Ardor	Dolor	0.960	0.8645833
## 1103	Ardor	Mal olor	0.680	0.7867647
## 1147	Irritación	Flujo vaginal	0.750	0.7700000
## 1454	Ureaplasma parvum Up	Ckg	0.845	0.7514793
## 1606	Sa	Alcohol	0.605	0.7107438
## 1744	AV alta densidad	Duchas mlvps	0.750	0.9833333

FIGURA 2. Resultados de la probabilidad condicional determinada como fuerza y dirección de los arcos en las variables cualitativas del dataset.

RESULTADOS

En los resultados finales se realizó la interpretación de los datos. Los resultados fueron dados con base en las experimentaciones de la sección experimentación de los datos anteriormente descrita.

Con respecto a los resultados mostrados en la sección de experimentación, se pudo percatar que todas las variables cualitativas se encuentran relacionadas sin tener una variable objetivo. Después de obtener las tablas de probabilidad de la BN con el algoritmo EM y mostrada en la Figura 2, se elaboraron las relaciones de probabilidad con base en la variable objetivo.

Como resultados finales se buscó encontrar las relaciones que tienen todas las variables cualitativas únicamente con la variable DxBV+. La relación de probabilidad resultante que contiene la relación de las variables cualitativas con la variable DxBV+ y construida con el algoritmo EM se muestra en la Figura 3.

En la Figura 3 se pueden ver la relación que tienen cada una de las variables cualitativas directamente con la variable objetivo. La variable objetivo se encuentra ahora en la columna *from*, las variables cualitativas están enlistadas en la columna *to* y en la columna *strength* y *direction* se cuenta con la probabilidad condicional representada entre estas dos variables.

Con las relaciones mostradas en la Figura 3 se interpreta que la variable AV alta densidad tiene una

##	from	to	strength	direction
## 1	Diagnóstico +BV No MH	Separacion trimestre	0.005	0.0000000
## 2	Diagnóstico +BV No MH	Diabetes	0.000	0.0000000
## 3	Diagnóstico +BV No MH	Hipertensión	0.005	0.5000000
## 4	Diagnóstico +BV No MH	Revisión ginecológica	0.005	0.0000000
## 5	Diagnóstico +BV No MH	Revisión nula	0.000	0.0000000
## 6	Diagnóstico +BV No MH	Una pareja	0.000	0.0000000
## 7	Diagnóstico +BV No MH	Dos parejas	0.000	0.0000000
## 8	Diagnóstico +BV No MH	Más de 3 parejas	0.000	0.0000000
## 9	Diagnóstico +BV No MH	Ácido fólico	0.000	0.0000000
## 10	Diagnóstico +BV No MH	Alcohol	0.000	0.0000000
## 11	Diagnóstico +BV No MH	Anticoncepción	0.080	0.5000000
## 12	Diagnóstico +BV No MH	Condón	0.000	0.0000000
## 13	Diagnóstico +BV No MH	Inyectado	0.015	0.1666667
## 14	Diagnóstico +BV No MH	Duchas Vaginales	0.025	0.0000000
## 15	Diagnóstico +BV No MH	Duchas mivps	0.010	0.0000000
## 16	Diagnóstico +BV No MH	Comezón genital	0.005	0.5000000
## 17	Diagnóstico +BV No MH	Ardor	0.130	0.4615385
## 18	Diagnóstico +BV No MH	Irritación	0.270	0.4722222
## 19	Diagnóstico +BV No MH	Dolor	0.020	0.1250000
## 20	Diagnóstico +BV No MH	Mal olor	0.005	0.0000000
## 21	Diagnóstico +BV No MH	Flujo vaginal	0.005	0.0000000
## 22	Diagnóstico +BV No MH	Ureaplasma parvum Up	0.010	0.5000000
## 23	Diagnóstico +BV No MH	Ca	0.475	0.3842105
## 24	Diagnóstico +BV No MH	Ckg	0.005	0.0000000
## 25	Diagnóstico +BV No MH	Sa	0.000	0.0000000
## 26	Diagnóstico +BV No MH	Lactobacillus Crispatus	0.040	0.2500000
## 27	Diagnóstico +BV No MH	Lactobacillus inners	0.005	0.0000000
## 28	Diagnóstico +BV No MH	AV alta densidad	0.970	0.6958763
## 29	Diagnóstico +BV No MH	GV alta densidad	0.010	0.2500000
## 30	Diagnóstico +BV No MH	MT1 alta densidad	0.965	0.7124352
## 31	Diagnóstico +BV No MH	BVAB2 alta densidad	0.745	0.6812081

FIGURA 3. Resultados de la probabilidad condicional determinada como fuerza y dirección de los arcos en la variable objetivo DxBV+ y las variables cualitativas del dataset.

probabilidad condicional del 97 %, MT1 alta densidad una probabilidad condicional de 96.5 %, BVAB2 alta densidad una probabilidad condicional de 74.5 % pero la bacteria GV alta densidad solo un porcentaje del 0.10 % aunque una relación del 25 % con respecto a las demás variables.

Para la creación del MGC se optó por tomar el algoritmo *Fast Causal Inference* FCI, ya que este algoritmo trabaja con variables discretas o categóricas lo que permitirá en trabajos futuros comparar con otros algoritmos que trabajen con datos cuantitativos. Se siguieron tomando las variables cualitativas.

Después de aplicar el algoritmo EM con los parámetros definidos por defecto, se fijaron los parámetros de búsqueda en un rango de probabilidad condicional mayor al 50 % con respecto a la variable objetivo DxBV+.

En la Figura 4, se usaron las 32 variables mostradas en los datos de la Figura 3, primero se realizó la estructura de la BN con el algoritmo EM y con el resultado obtenido se aplicó el algoritmo FCI.

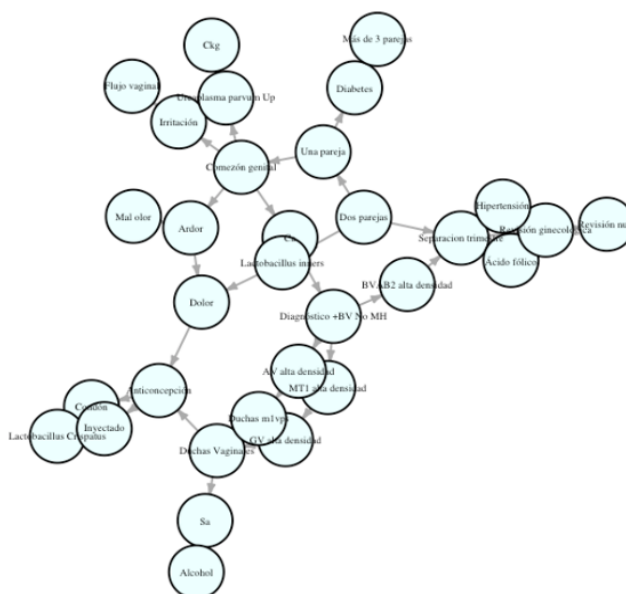


FIGURA 4. Gráfico de BN con algoritmo EM y algoritmo FCI con nodos y arcos dirigidos entre las variables cualitativas y la variable objetivo $DxBV+$.

En la Figura 4 se pueden apreciar nuevamente las relaciones dadas entre las variables cualitativas mediante nodos y arcos. Cuenta con nodos (aristas) y flechas (arcos dirigidos). En esta ocasión se toma como el nodo (A) a la variable Comezón genital y a la variable Ardor como el nodo (B), se interpreta que el nodo A tiene un efecto causal o una influencia directa sobre el nodo B.

Hasta la Figura 4 se tienen relacionadas 32 variables cualitativas que tienen relación directa con el DxBV+, en la siguiente delimitación se reducen las variables que tienen un mayor porcentaje de relación.

El porcentaje de relación para el ejercicio es mayor al 50 % ya que el objetivo es encontrar las variables que tengan una mayor relación con el DxBV+.

De 32 variables iniciales, la experimentación concluye con 22 variables cualitativas incluyendo la variable objetivo. Las 22 variables se encuentran representadas en la BN mostrada en la Figura 5.

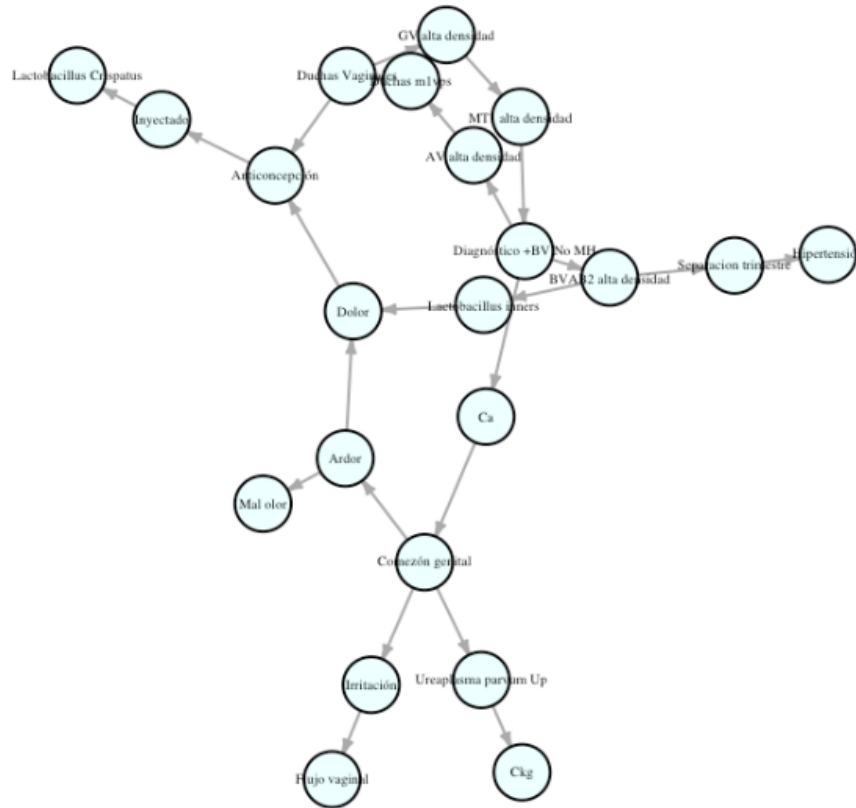


FIGURA 5. Red Bayesiana Causal CBN creada con las 22 variables que tienen relación con el Diagnóstico BV+.

Interpretación y discusión de los resultados

La Figura 5 presenta la CBN en la se tiene la presencia de 22 variables cualitativas que tienen una relación directa con el diagnóstico positivo de la Vaginosis Bacteriana. En la CBN de la Figura 5 se puede apreciar la distribución de los nodos y arcos.

En esta ocasión a manera de ejemplo, se toma como el nodo (A) a la variable MT1 alta densidad y a la variable Diagnóstico +BV para el nodo (B), se interpreta que el nodo A tiene un efecto causal o una influencia directa sobre el nodo B. Otro ejemplo, se puede apreciar en la variable BVAB2 alta densidad nodo (A) con respecto a la variable Lactobacillus inners nodo (B) y también se puede interpretar que la variable BVAV2 alta densidad tiene una influencia directa o un efecto causal sobre la variable

Lactobacillus inners.

La CBN creada se puede seguir explicando nodo por nodo, sin embargo, también es importante mencionar que esta CBN tiene como resultados relaciones de probabilidad condicional. Las relaciones de probabilidad condicional de la CBN muestran el porcentaje de relaciones existentes entre las variables cualitativas que tienen un porcentaje mayor al 50 % con respecto a la variable objetivo Diagnóstico BV+.

##		from	to	strength	direction
## 22	Separacion trimestre	Diagnóstico +BV No MH	0.000	0.0000000	
## 43	Hipertensión	Diagnóstico +BV No MH	0.005	0.5000000	
## 64	Anticoncepción	Diagnóstico +BV No MH	0.120	0.6666667	
## 85	Inyectado	Diagnóstico +BV No MH	0.025	0.7000000	
## 106	Duchas Vaginales	Diagnóstico +BV No MH	0.025	0.7000000	
## 127	Duchas mlvps	Diagnóstico +BV No MH	0.000	0.0000000	
## 148	Comezón genital	Diagnóstico +BV No MH	0.005	0.0000000	
## 169	Ardor	Diagnóstico +BV No MH	0.175	0.5428571	
## 190	Irritación	Diagnóstico +BV No MH	0.310	0.5322581	
## 211	Dolor	Diagnóstico +BV No MH	0.015	0.8333333	
## 232	Mal olor	Diagnóstico +BV No MH	0.000	0.0000000	
## 253	Flujo vaginal	Diagnóstico +BV No MH	0.005	0.5000000	
## 274	Ureaplasma parvum Up	Diagnóstico +BV No MH	0.020	0.5000000	
## 295	Ca	Diagnóstico +BV No MH	0.535	0.5560748	
## 316	Ckg	Diagnóstico +BV No MH	0.000	0.0000000	
## 337	Lactobacillus Crispatus	Diagnóstico +BV No MH	0.040	0.8750000	
## 358	Lactobacillus inners	Diagnóstico +BV No MH	0.000	0.0000000	
## 379	AV alta densidad	Diagnóstico +BV No MH	0.980	0.3520408	
## 400	GV alta densidad	Diagnóstico +BV No MH	0.035	0.7857143	
## 421	MT1 alta densidad	Diagnóstico +BV No MH	0.935	0.3449198	
## 442	BVAB2 alta densidad	Diagnóstico +BV No MH	0.680	0.3161765	

FIGURA 6. Resultados de la probabilidad condicional determinada como fuerza y dirección de los arcos en la variable objetivo DxBV+ y las variables cualitativas finales del dataset.

Finalmente, en la Figura 6 se presentan las relaciones de probabilidad pero de manera inversa a como se presentaron en la Figura 2 y Figura 3. En la columna descrita como *to*, ahora se encuentra la variable objetivo DxBV+, en *from* la variable cualitativa relacionada y finalmente *strength* y *direction* detallan el porcentaje de probabilidad condicional que tienen en relación estas dos variables. Las relaciones de probabilidad se pueden apreciar que las variables mayores a 60 % son AV alta densidad, MT1 alta densidad y BVAB2 alta densidad que son bacterias que de acuerdo con la literatura están relacionadas con el Diagnóstico BV+ [14].

CONCLUSIONES

De acuerdo con la implementación del algoritmo de Red Bayesiana Causal en el *dataset* de Vaginosis Bacteriana en mujeres embarazadas, las variables que se obtuvieron y tienen una relación mayor al 50 % con respecto a la variable del Diagnóstico BV+ fueron:

Anticoncepción, Inyectado, Duchas vaginales, Ardor, Irritación, Dolor, Ca, Lactobacillus crispatus, GV alta densidad.

Por otro lado, las variables que tienen un porcentaje mayor al 60 % en fuerza de correspondencia con respecto a la variable Diagnóstico BV+ fueron:

AV alta densidad, MT1 alta densidad y BVAB2 alta densidad.

Las variables con porcentaje mayor al 60 % coinciden con las que hay en la literatura. El presente estudio muestra que hay bacterias que coinciden con el DxBV+. Algunos estudios señalan que las variables como las duchas vaginales, inyecciones anticonceptivas y calcio también pueden influir en un mínimo porcentaje en el desarrollo de la enfermedad conocida como Vaginosis Bacteriana [15] .

TRABAJOS FUTUROS

Como trabajo futuro se tiene el objetivo de realizar experimentos con otros algoritmos como *tabu search*, *hill climbing* para evaluar el desempeño del algoritmo CBN.

Así mismo se pretende realizar un estudio con variables cuantitativas donde se evalúe el desempeño de Redes Bayesianas Causales desde la perspectiva de los datos cuantitativos del *dataset*.

REFERENCIAS

- [1] J. B. Mina-Ortiz, M. O. Franco-Macias, L. A. Santana, and M. G. Garcia-Ortega. Impacto en la salud materno fetal de embarazadas adolescentes con vaginosis bacteriana. *MQRInvestigar*, 8(1), pp. 5041–5264, 2024.
- [2] Z. Rivera, N. Bravo, R. Borges, and M. Ferreira. Infecciones de transmisión sexual y embarazo. 2010.
- [3] A. A. Lachira Vasquez. Factores asociados a vaginosis bacteriana en mujeres en edad reproductiva en una localidad rural y otra urbano-marginal de Piura–2021. 2023.
- [4] G. Perez Coronel. Conocimiento y actitudes sobre prevención de infecciones de transmisión sexual en estudiantes del Colegio Sagrado Corazón de Jesús, Chota 2023. 2024.
- [5] S. Swidsinski, W. M. Moll, and A. Swidsinski. Bacterial Vaginosis-Vaginal Polymicrobial Biofilms and Dysbiosis. *Deutsches Ärzteblatt International*, 120(20), p. 347, 2023.
- [6] M. Á. Vega, L. M. Q. Mora, and M. V. C. Badilla. Inteligencia artificial y aprendizaje automático en medicina. *Revista médica sinergia*, 5(8), pp. e557–e557, 2020.
- [7] F. Cantú and R. Aldeco. Conocimiento y Razonamiento Computacional. 2019.

- [8] E. M. Rojas. Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E28), pp. 586–599, 2020.
- [9] G. Soto. El teorema de Bayes. *Revista de Educación Matemática*, 26(3), 2011.
- [10] J. Pearl and D. Mackenzie. El libro del porqué. *La nueva ciencia de la causa y el efecto*. Barcelona, España: Pasado y Presente, 2020.
- [11] L. E. Sucar. Probabilistic graphical models. *Advances in Computer Vision and Pattern Recognition*. London: Springer London, 2015. DOI: 10.1007/978-1-4471-6692-5
- [12] M. Lajous. Inferencia causal en análisis basados en datos de vigilancia epidemiológica para Covid-19. *Salud pública de méxico*, 63(4), pp. 459–460, 2021.
- [13] K. Train. Métodos de elección discreta con simulación. 2014.
- [14] C. E. Yucra López. Biblioteca para la comparación estadística de algoritmos evolutivos. 2021.
- [15] K. A. M. Morón and C. A. M. Romero. Prevalencia y Factores Asociados con Vaginosis Bacteriana en Universitarias de la Medicina. *Medicina*, vol. 15, p. 15.