

ANÁLISIS CLÚSTER DE INFORMACIÓN SOBRE INFECCIONES URINARIAS

RUTH MARÍA REÁTEGUI ROJAS

rmreategui@utpl.edu.ec

<https://orcid.org/0000-0002-7304-4413>

Universidad Técnica Particular de Loja, Ecuador

MARÍA IRENE CARRILLO MAYANQUER

micarrillo@utpl.edu.ec

<https://orcid.org/0000-0001-9968-2539>

Universidad Técnica Particular de Loja, Ecuador

Recibido: 2 de agosto del 2024 / Aceptado: 14 de octubre del 2024

doi: <https://doi.org/10.26439/interfases2024.n020.7327>

RESUMEN. Las infecciones urinarias constituyen el principal motivo de consulta en el servicio de urgencias pediátricas en el mundo, por lo que merecen ser analizadas con técnicas de inteligencia artificial que permitan descubrir patrones basados en información médica y de laboratorio. El análisis clúster es una técnica no supervisada de aprendizaje de máquina que permite identificar grupos de pacientes con características similares. En este trabajo, se analizó información anonimizada de pacientes extraída de un sistema informático, donde todos sufren de infecciones urinarias. Se aplicó inicialmente el análisis de correspondencia múltiple (ACM) para luego utilizar de forma separada los algoritmos K-means y DBSCAN. Se obtuvo el valor de *silhouette* de cada grupo obtenido con los dos algoritmos. Se logró diferenciar a los pacientes de acuerdo con los porcentajes de prevalencia de sensibilidad/resistencia a ciertos antibióticos y a la presencia de los gérmenes que provocan las infecciones.

PALABRAS CLAVE: inteligencia artificial / aprendizaje de máquina / salud

CLUSTER ANALYSIS OF INFORMATION ON URINARY TRACT INFECTIONS

ABSTRACT. Urinary tract infections are the main reason for consultation in the pediatric emergency department worldwide, so it deserves to be analyzed with artificial intelligence techniques to discover patterns based on medical and laboratory information. Cluster analysis is an unsupervised machine learning technique that allows the identification of groups of patients with similar characteristics. In this work we analyzed information from patients whose anonymized information was extracted from a computer system, all of them are patients suffering from urinary tract infections.

Multiple Correspondence Analysis was initially applied and then K-means and DBSCAN algorithms were used separately. The silhouette value of each group identified with the two algorithms was obtained. Patients were differentiated according to the prevalence percentages of sensitivity/resistance to certain antibiotics and the presence of the germs causing the infections.

KEYWORDS: artificial intelligence / machine learning / health

INTRODUCCIÓN

La infección del tracto urinario (ITU) se produce por la colonización de bacterias en esa zona (Hevia et al., 2020). Este es el principal motivo de consulta en el servicio de urgencia pediátrica a nivel mundial, pues representa el 14 % de visitas al año (Ardila et al., 2015). Además, es una de las principales causas de fiebre sin foco en lactantes (Leung et al., 2019). La ITU tiene una prevalencia general de aproximadamente el 7 % en los lactantes que presentan fiebre, pero esta prevalencia varía según la edad, el sexo y el estado de circuncisión. Las niñas tienen una prevalencia de infecciones urinarias entre dos y cuatro veces mayor que los hombres circuncidados; sin embargo, esta prevalencia se iguala en los niños no circuncidados, especialmente en menores de 3 meses (Shaikh & Hoberman, 2022). La ITU alta o pielonefritis a repetición puede causar cicatrización renal, hipertensión arterial y enfermedad renal crónica, de allí que un diagnóstico precoz y un tratamiento adecuado son de importancia para prevenir complicaciones a futuro (Shaikh & Hoberman, 2022).

Los microorganismos más comunes provienen de la flora intestinal. Por ejemplo, la *Escherichia coli* es responsable del 80 % al 90 % de las infecciones urinarias en niños (Leung et al., 2019). En los últimos años, se ha detectado una disminución progresiva en la sensibilidad de este microorganismo frente a los antimicrobianos (Betrán et al., 2020). En particular, los mecanismos subyacentes de virulencia y resistencia a los antibióticos en niños no son bien conocidos (Salazade et al., 2023). Según Cavagnaro Santa María (2014), la resistencia antibiótica frecuentemente está relacionada con el uso prolongado y, a veces, inapropiado de estos fármacos, de modo que considera necesaria una evaluación periódica de la prevalencia y la resistencia bacteriana de estos microorganismos.

Actualmente, con mayor fuerza se busca implementar la inteligencia artificial (IA) en varios campos del conocimiento, entre los cuales la salud es uno de los más destacados. El análisis clúster es una de las técnicas de IA que permite descubrir patrones y segmentar a los pacientes en grupos homogéneos, basados en características como las demográficas, clínicas y de laboratorio. Además, según la información proveniente de sistemas informáticos médicos, son varios trabajos los que se han desarrollado con análisis clúster. Por ejemplo, en análisis de enfermedades relacionadas con el corazón, Martins et al. (2014), Wazzan et al. (2024), Banerjee et al. (2023) y Hobensack et al. (2023) aplicaron algoritmos jerárquicos, K-medias y K-modas. Otros trabajos, que utilizaron los mismos algoritmos, se han enfocado en analizar información de la columna vertebral (Mohanty et al., 2024), de la lesión renal (Bolt et al., 2023) y de la epilepsia (Josephson et al., 2023). Además, Chacaguasay et al. (2024) analizaron infecciones urinarias en mujeres y compararon los resultados con tres algoritmos: K-means, DBSCAN y jerárquico. La desnutrición en niños también ha sido estudiada por Hadikurniawati et al. (2023), estos autores aplicaron K-means y DBSCAN. También, varios de estos trabajos aplicaron técnicas para la reducción de la dimensionalidad y algunas métricas para el cálculo de la similitud como Gower y Jaccard.

Entre los trabajos más cercanos al análisis de resistencia/sensibilidad a antibióticos e incluso al análisis de las bacterias que causan enfermedades en niños, se encuentra el trabajo de Krueger et al. (2024), donde utilizaron K-means para analizar 10 609 prescripciones de antibióticos en tres diferentes enfermedades: infección del tracto urinario, neumonía adquirida en la comunidad y otitis media aguda. El resultado arrojó dos grupos de pacientes, cuya diferencia radica en el tiempo de prescripción de los antibióticos. Las mismas enfermedades fueron analizadas por Elgoibar et al. (2020) en 291 niños infectados con *E. coli*, donde identificaron cuatro grupos de pacientes claramente diferenciados por su condición médica. Por otro lado, Suwono et al. (2021) utilizaron información de la resistencia a los antibióticos en *E. coli* procedentes de diferentes sistemas de vigilancia y monitoreo para humanos y diferentes poblaciones animales. Utilizaron el algoritmo jerárquico e identificaron varios clústeres diferenciados por las combinaciones de resistencia a los antibióticos.

Por lo antes expuesto, el análisis clúster es una técnica que no solo facilita la comprensión de la variabilidad de las enfermedades, sino que también puede informar estrategias de tratamiento personalizadas y la gestión de la enfermedad. Por tal razón, el presente estudio se enfoca en aplicar K-means y DBSCAN, algoritmos ampliamente utilizados en el ámbito médico, como se ha demostrado en los párrafos anteriores, para explorar y caracterizar subgrupos de niñas con infecciones urinarias. Se ha utilizado un conjunto de datos que incluye variables clínicas, demográficas y de laboratorio.

En el presente trabajo, primero se brinda una introducción al tema a tratar; luego, se describe la metodología utilizada; y, finalmente, se ofrecen los resultados y las conclusiones.

1. METODOLOGÍA

En este apartado, se explicarán las fases que se llevaron a cabo para el proceso de experimentación.

1.1 Dataset

Para el presente estudio, se utilizó una base de 133 niñas cuyo diagnóstico ha sido codificado como "infección de vías urinarias en sitio no especificado". Esta base incluye información sobre la sensibilidad y resistencia a ciertos antibióticos, así como datos de los gérmenes identificados en los exámenes de laboratorio.

1.2 Preprocesamiento

Las variables fueron configuradas para que el valor 1 indique la presencia de sensibilidad o resistencia a determinado antibiótico, y 0 para el caso contrario. Además, la variable germen fue codificada para que cada patógeno sea una variable; es decir, se

crearon variables *dummy*, donde el valor 1 corresponde a la presencia de un germen específico y 0 para el caso contrario.

Se procedió a crear las variables *SOtro* (abreviación para sensibilidad_otros), *ROtro* (abreviación para resistencia_otros) y *GermenOtro* para agrupar las variables de sensibilidad, resistencia y gérmenes que estaban presentes en uno o dos pacientes. Por ejemplo, de las pacientes analizadas solo una presentó sensibilidad a amoxicilina + ácido clavulánico, por ende, esta variable se añadió a *SOtro*. Con esto, se tuvo un corpus de 133 pacientes y 48 variables dicotómicas, de las cuales 24 están relacionadas a sensibilidad, 19 a resistencia y 5 a gérmenes. La Tabla 1 muestra el listado de las 48 variables.

1.3 Análisis clúster

Para el proceso de experimentación, se trabajó con los algoritmos no supervisados K-means y DBSCAN. Previo a la utilización de estos algoritmos, se realizó la reducción de variables mediante el método MCA. A continuación, se brinda una breve explicación de las técnicas utilizadas y la justificación de su uso en este trabajo.

MCA

Existen varias técnicas de reducción de dimensionalidad, uno de los más populares es el análisis de componentes principales (PCA) que trabaja con datos numéricos. Nuestro *dataset* contiene variables categóricas (binarias), por lo que se requiere una técnica que permita tratar este tipo de dato. El análisis de correspondencia múltiple (MCA, por sus siglas en inglés) es una técnica de reducción de dimensionalidad que trabaja con datos categóricos. Para la reducción de la dimensionalidad, se selecciona un número determinado de componentes principales (eigenvectores) que explican la mayor parte de la variabilidad en los datos, similar a como trabaja el PCA. Así, el MCA captura la mayor parte de la variabilidad en los datos categóricos, lo que permite que los algoritmos de agrupamiento operen de manera más eficiente y efectiva. Es importante recalcar que luego de aplicar este método se obtendrá una matriz con dos componentes y con valores numéricos, por tanto, los algoritmos para el análisis clúster deben ser aptos para *data-sets* numéricos.

K-means

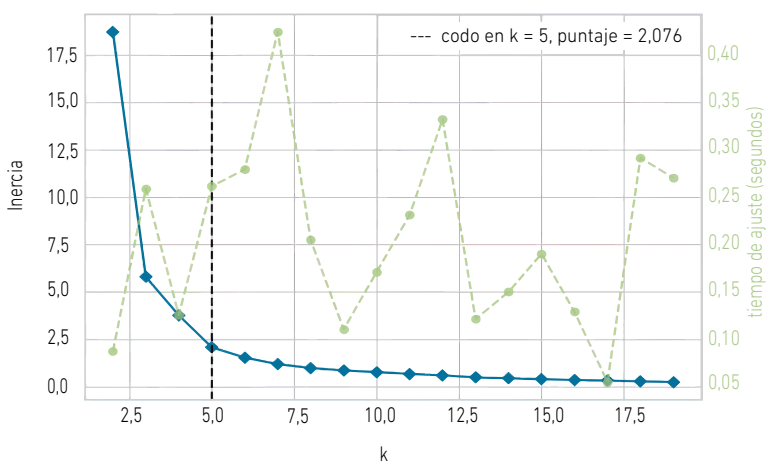
El algoritmo K-means es un método no supervisado que se utiliza para dividir un conjunto de datos en un número predeterminado de clústeres (k) basados en similitudes en las características de los datos. La elección de este algoritmo se basa en que, aparte de ser ampliamente utilizado en el ámbito médico, es un algoritmo simple, computacionalmente eficiente y que converge rápidamente hacia un resultado, lo cual es beneficioso para análisis preliminares o cuando se requiere celeridad en los resultados.

Método del codo

Este método se basa en calcular y visualizar los valores de la inercia obtenidos al calcular el algoritmo K-means para un rango determinado de valores de k . Se grafica la inercia en función del número de clústeres k ; en el eje x se coloca el número de clústeres y en el eje y , la inercia calculada. Aplicando este método, el k recomendado para trabajar con el algoritmo K-means es de cinco grupos; esto se puede observar en la Figura 1.

Figura 1

Método del codo para determinar el número de clústeres (k)



Nota. En la gráfica, se puede observar que el valor del codo en k es igual a 5.

DBSCAN

Este algoritmo DBSCAN (siglas en inglés de density-based spatial clustering of applications with noise) utiliza la densidad para la identificación de los clústeres o grupos. Trabaja con dos parámetros: ϵ (eps), que es la distancia máxima para que dos objetos sean reconocidos como vecinos, y el número mínimo de puntos (minPts) requeridos para que una región se considere densa. Este algoritmo también ha sido ampliamente utilizado en datos médicos. A diferencia de K-means, este algoritmo no necesita previamente conocer el número de clústeres a identificar. También es importante recalcar que DBSCAN puede identificar valores atípicos, la cual es una de sus fortalezas con respecto a K-means.

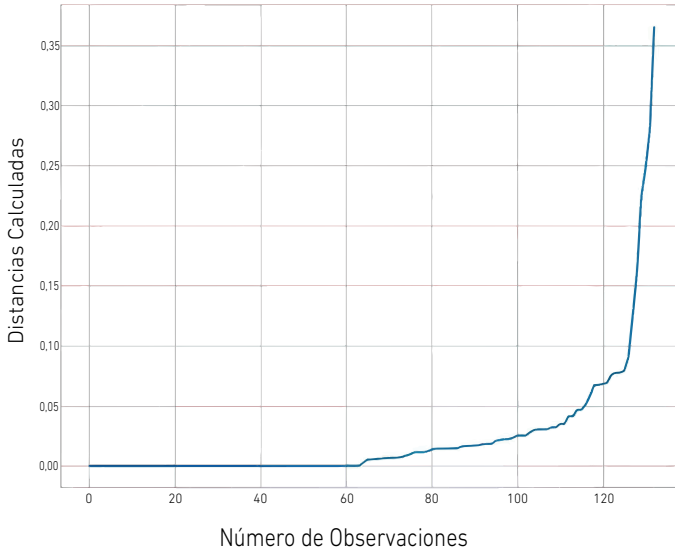
Gráfico de distancia basado en k vecinos más cercanos (KNN)

Siendo ϵ uno de los parámetros relevantes para DBSCAN, se puede recurrir al gráfico de distancia basados en vecinos para identificar un valor aproximado de ϵ . En la

Figura 2, se puede observar el gráfico de las distancias calculado para cinco vecinos, el valor que tomaremos es de 0,06.

Figura 2

Gráfica de distancias para cinco vecinos



2. DISCUSIÓN DE RESULTADOS

Luego de aplicar K-means con un k igual a 5, se obtuvieron los resultados que se muestran en la Tabla 1; se puede observar los cinco grupos nombrados de 0 a 4. Cada uno de los grupos tiene 52, 36, 5, 35 y 5 pacientes, respectivamente.

Tabla 1

Resultados del análisis clúster con K-means

Clústeres	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Número de pacientes	52	36	5	35	5
<i>Silhouette</i>	0,35	0,87	0,81	0,66	0,43
SENSIBILIDAD					
Amikacina	62	0	80	100	80
Fosfomicina	54	0	40	86	20
Nitrofurantóina	60	3	100	89	20
Gentamicina	69	3	40	89	60
Trimetoprima-sulfametoxazol	35	0	40	40	20
Ciprofloxacina	17	0	20	26	20

(continúa)

(continuación)

Clústeres	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo 4
SENSIBILIDAD					
Cefuroxima	62	0	0	97	20
Ampicilina + sulbactam	42	0	0	37	20
Cefepima	4	0	0	34	40
Ampicilina	10	0	0	11	20
Norfloxacin	29	0	0	3	0
Ácido Nalidíxico	21	0	0	14	0
Ceftazidima	17	0	0	34	20
Cefazolina	12	0	0	97	0
Cefalexina	15	0	0	63	0
Cefotaxima	31	0	0	60	20
Amoxicilina + ácido clavulánico	19	3	0	40	0
Ceftriaxona	56	0	0	34	20
Levofloxacin	6	3	0	3	20
Ertapenem	0	0	100	0	60
Meropenem	0	0	40	0	100
Imipenem	0	0	80	0	80
Piperacilina-tazobactam	0	0	80	0	60
S0tro	8	3	0	0	0
RESISTENCIA					
Trimetoprima-sulfametoxazol	46	3	20	49	80
Amoxicilina + ácido clavulánico	13	0	20	23	20
Fosfomicina	19	0	40	6	40
Ciprofloxacina	4	0	80	9	40
Ampicilina + sulbactam	10	0	100	3	40
Gentamicina	8	0	0	6	40
Ampicilina	63	0	0	49	20
Cefalexina	15	3	0	0	0
Nitrofurantoína	15	3	0	3	0
Norfloxacin	6	0	20	3	0
Aztreonam	0	0	100	0	0
Ceftriaxona	0	0	100	0	80
Cefepima	0	0	100	0	20
Ceftazidima	2	0	100	0	40
Cefotaxima	0	0	100	0	60
Cefazolina	13	0	100	0	20
Cefuroxima	4	0	100	0	80

(continúa)

(continuación)

Clústeres	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo 4
RESISTENCIA					
Ácido nalidíxico	12	0	80	0	0
ROtro	17	0	0	6	60
GERMEN					
Germen_ <i>E. Coli</i>	81	6	0	94	40
Germen_ <i>E. Coli BLEE</i>	0	0	80	0	0
Germen_ <i>Enterobacter cloacae</i>	6	0	0	6	0
Germen_ Flora mixta	0	25	0	0	0
Germen Otro	12	3	0	0	20

Nota. Los resultados incluyen la cantidad de pacientes por clúster y los porcentajes de pacientes que presentan sensibilidad/resistencia a los antibióticos o gérmenes.

En todos los grupos, excepto el grupo 1, se encontraron pacientes con sensibilidad a antibióticos como amikacina, fosfomicina, nitrofurantoína, gentamicina, trimetoprima-sulfametoxazol, ciprofloxacina, y resistencia a trimetoprima-sulfametoxazol, amoxicilina + ácido clavulánico, fosfomicina, ciprofloxacina y ampicilina + sulbactam. Hay ciertos antibióticos que sobresalen por el porcentaje de pacientes que presentan resistencia o sensibilidad. A continuación, se brinda mayor detalle de los grupos encontrados.

El grupo 1 está conformado por 36 pacientes. Este grupo se destaca porque el 25 % de las pacientes presenta flora mixta. Respecto a la sensibilidad y resistencia a antibióticos, no hay mayor relevancia de la cantidad de pacientes que presentan este tipo de condiciones.

El grupo 2 está compuesto por cinco pacientes, de los cuales el 80 % (4) presenta la bacteria *E. Coli BLEE*. Además de la resistencia y sensibilidad a los antibióticos comunes en todos los grupos, este grupo tiene un alto porcentaje de pacientes con sensibilidad al imipenem (80 %), piperacilina-tazobactam (80 %) y ertapenem (100 %). Asimismo, hay un alto porcentaje de resistencia al ácido nalidíxico (80 %) y todas las pacientes son resistentes a la ceftriaxona, cefepima, ceftazidima, cefotaxima, cefazolina, cefuroxima y aztreonam.

El grupo 4 está conformado por cinco pacientes que no presentan algún germen en especial. Este grupo tiene pocos pacientes (uno o dos) con sensibilidad y resistencia a varios antibióticos; sin embargo, al igual que el grupo 2, el 80 % de las pacientes muestra sensibilidad a imipenem. Además, todas son sensibles a meropenem. En cuanto a la resistencia, el 80 % es resistente a ceftriaxona y cefuroxima.

Los grupos 0 y 3 tienen 52 y 35 pacientes cada uno, respectivamente. La mayoría presenta el germen *E. Coli*, el 81 % en el grupo 0 y el 94 % en el grupo 3. Estos grupos son diferentes a los grupos 2 y 4 por la presencia de pacientes con resistencia a ciertos

antibióticos. Los grupos 0 y 3 presentan un 63 % y 49 % de pacientes, respectivamente, con resistencia a la ampicilina. El grupo 3 se destaca por contar con un mayor porcentaje de pacientes con sensibilidad a la amoxicilina + ácido clavulánico (40 %), cefazolina (97 %), cefalexina (63 %) y cefotaxima (60 %). Por el contrario, el grupo 0 posee el mayor porcentaje de pacientes con sensibilidad a la ceftriaxona (56 %). Ambos grupos también presentan alto porcentaje de pacientes con sensibilidad a la cefuroxima: 62 % en el grupo 0 y 97 % en el grupo 3.

Se realizó otro experimento con el algoritmo DBSCAN, para el cual se eligió trabajar con una eps de 0,06 y minPts de 5, valores tomados de acuerdo con la Figura 2. Los resultados se muestran en la Tabla 2. Los cuatro grupos identificados son nombrados de 0 a 3; además, por motivos de comparación con el experimento anterior, en la Tabla 2 se decidió ubicar el grupo 1, que corresponde a los valores atípicos. Cada uno de los grupos tiene 62, 7, 24, 9 y 31 pacientes, respectivamente. Este nuevo experimento separa notablemente, en los grupos 2 y 3, a las pacientes que no presentan ninguna de las condiciones consideradas como variables de análisis. La gran diferencia entre estos grupos es que, en el grupo 3, las nueve personas que forman parte de este grupo cuentan con el Germen_Flora mixta.

Además, podemos observar que en este experimento se pierden los clústeres que permiten identificar a las pacientes con resistencia a ciertos antibióticos, como son los casos de los grupos 2 y 4 de K-means. De este modo, quienes los conformaban ahora fueron identificadas como atípicas en el grupo -1 de DBSCAN.

En los grupos 0 y 1, se hallan las pacientes con características muy similares respecto a la sensibilidad y resistencia a ciertos antibióticos; la diferencia entre ellos radica en los porcentajes de pacientes que presentan cierta condición. Por ejemplo, el grupo 0 cuenta con porcentajes sobre 69 para la sensibilidad a la amikacina, fosfomicina, nitrofurantoína y gentamicina, y para el grupo 1 los porcentajes son menores a 43.

Tabla 2

Resultados del análisis clúster con DBSCAN

Clústeres	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo -1
Número de pacientes	62	7	24	9	31
<i>Silhouette</i>	0,33	0,78	0,99	1,00	
	SENSIBILIDAD				
Amikacina	89	14	0	0	61
Fosfomicina	69	14	0	0	55
Nitrofurantoína	82	43	0	0	48
Gentamicina	87	43	0	0	52
Trimetoprima-sulfametoxazol	39	71	0	0	19

(continúa)

(continuación)

Clústeres	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo -1
SENSIBILIDAD					
Ciprofloxacina	26	0	0	0	13
Cefuroxima	84	100	0	0	26
Ampicilina + sulbactam	34	71	0	0	32
Cefepima	19	0	0	0	13
Ampicilina	10	43	0	0	3
Norfloxacina	21	0	0	0	10
Ácido Nalidíxico	23	29	0	0	0
Ceftazidima	23	0	0	0	26
Cefazolina	56	14	0	0	13
Cefalexina	42	29	0	0	6
Cefotaxima	40	71	0	0	26
Amoxicilina + ácido clavulánico	34	14	0	0	10
Ceftriaxona	40	57	0	0	42
Levofloxacina	3	0	0	0	13
Ertapenem	0	0	0	0	26
Meropenem	0	0	0	0	23
Imipenem	0	0	0	0	26
Piperacilina-tazobactam	0	0	0	0	23
RESISTENCIA					
Trimetoprima-sulfametoxazol	44	14	0	0	61
Amoxicilina + ácido clavulánico	18	0	0	0	19
Fosfomicina	15	14	0	0	19
Ciprofloxacina	5	0	0	0	26
Ampicilina + sulbactam	2	0	0	0	39
Gentamicina	5	14	0	0	13
Ampicilina	58	14	0	0	45
Cefalexina	8	14	0	0	10
Nitrofurantoína	10	14	0	0	10
Norfloxacina	0	0	0	0	16
Aztreonam	0	0	0	0	16
Ceftriaxona	0	0	0	0	29
Cefepima	0	0	0	0	19
Ceftazidima	0	0	0	0	26
Cefotaxima	0	0	0	0	26

(continúa)

(continuación)

Clústeres	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo -1
RESISTENCIA					
Cefazolina	6	0	0	0	29
Cefuroxima	0	0	0	0	35
Ácido nalidíxico	3	0	0	0	26
GERMEN					
Germen_ <i>E. Coli</i>	90	71	0	0	58
Germen_ <i>E. Coli BLEE</i>	0	0	0	0	13
Germen_ <i>Enterobacter cloacae</i>	6	14	0	0	0
Germen_Flora mixta	0	0	0	100	0
R0tro	8	0	0	0	29
S0tro	5	0	0	0	6
GermenOtro	2	14	0	0	19

3. CONCLUSIONES

El análisis clúster permitió identificar varios subgrupos clínicos entre las niñas con infecciones urinarias, en donde el germen presente con mayor frecuencia fue la *E. coli*; cada uno presentaba características de sensibilidad y resistencia diferentes a los antibióticos. Aquí se puede destacar uno de los grupos K-means (grupo 2), donde se observa al germen *E. coli*, productora de betalactamasa de ESPECTRO extendido (BLEE). A pesar de ser un grupo pequeño, se distingue a causa de la gran cantidad de antibióticos a los cuales las niñas son resistentes, como son todas las cefalosporinas (ceftriaxona, cefepima, ceftazidima, cefotaxima, cefazolina y cefuroxima). El tratamiento indicado de acuerdo con el resultado del antibiograma son los carbapenémicos (imipenem y ertapenem) y la piperacilina-tazobactam. Toda esta información se refleja en el grupo identificado por el algoritmo.

Los resultados son una evidencia de lo que está ocurriendo a nivel internacional sobre la resistencia bacteriana a los antibióticos comúnmente utilizados en el tratamiento empírico en las infecciones urinarias, así como de la generación de cepas productoras de betalactamasa de espectro extendido (BLEE). Con esto se puede facilitar la toma de decisiones en el tratamiento de infecciones urinarias de acuerdo con la evidencia y la epidemiología local, la misma que hasta el momento no se dispone. Por lo tanto, este estudio es un aporte para el tratamiento de infecciones urinarias en la región sur del Ecuador.

Los resultados de los algoritmos presentaron algunas semejanzas en ciertos grupos. El grupo 0, de ambos experimentos, agrupa a pacientes con sensibilidad y resistencia a ciertos antibióticos. El grupo 1 de K-means y 2 de DBSCAN son muy semejantes al presentar porcentajes bajos de pacientes en algunas de las condiciones de análisis. Si bien los dos algoritmos se caracterizan por contar con formas diferentes de identificar

los grupos, K-means se basa en distancias y DBSCAN, en densidad; en este experimento, DBSCAN tuvo valores más altos para *silhouette* en cada uno de los grupos identificados. Este algoritmo se destaca por identificar grupos más limpios, en el sentido de que todos los pacientes de un grupo comparten las mismas características, como es el caso de los grupos 2 y 3. En el grupo 2, ninguna de las pacientes presenta alguna condición de las analizadas. En el grupo 3, todas las pacientes presentan una única condición, que es el Germen_Flora mixta. Si bien este algoritmo no identifica a un grupo de pacientes con resistencia a ciertos antibióticos, DBSCAN los reconoce como atípicos.

Una de las principales limitantes de este trabajo fue la cantidad de registros a los que se tuvo acceso. Igualmente, solo se trabajó con información de los urocultivos con resultados infecciosos, el cual brindó la información del tipo de germen, la resistencia y sensibilidad antimicrobiana. Con respecto a la edad, se trabajó con niñas de 3 meses a 5 años; esta variable fue eliminada, ya que el rango de edad no se consideró relevante para este experimento. Asimismo, variables como lugar de nacimiento o residencia no fueron entregadas por cuestiones de anonimato; no obstante, cabe recalcar que, al tratarse de una muestra pequeña de un hospital local, las pacientes generalmente provienen del mismo lugar o lugares cercanos. Los resultados fueron obtenidos de una data de diez años, que podría justificar la variabilidad de la resistencia y la sensibilidad antimicrobiana, como se puede observar en otros estudios realizados (Whelan et al., 2024). Con esta limitante, también se presenta una oportunidad para que en un trabajo futuro se considere ampliar el número de registros y obtener otro tipo de información tanto clínica como demográfica.

A nivel técnico, para próximas investigaciones se podría ampliar la cantidad de algoritmos a utilizar o incluso hacer una combinación de los ya utilizados en este proyecto. También se podría comparar los resultados con otras técnicas de reducción de dimensionalidad, como el PCA logístico.

REFERENCIAS

- Ardila, M., Rojas, M., Santisteban, G., Gamero, A., & Torres, A. (2015). Infección urinaria en pediatría. *Repertorio de Medicina y Cirugía*, 24(2), 113-122. <https://doi.org/10.31260/RepertMedCir.v24.n2.2015.632>
- Banerjee, A., Dashtban, A., Chen, S., Pasea, L., Thygesen, J. H., Fatemifar, G., Tyl, B., Dyszynski, T., Asselbergs, F. W., Lund, L. H., Lumbers, T., Denaxas, S., & Hemingway, H. (2023). Identifying subtypes of heart failure from three electronic health record sources with machine learning: An external, prognostic, and genetic validation study. *The Lancet Digital Health*, 5(6), e370-e379. [https://doi.org/10.1016/s2589-7500\(23\)00065-1](https://doi.org/10.1016/s2589-7500(23)00065-1)
- Betrán, A., Lavilla, M. J., Cebollada, R., Calderón, J. M., & Torres, L. (2020). Resistencia antibiótica de *Escherichia coli* en infecciones urinarias nosocomiales y

- adquiridas en la comunidad del Sector Sanitario de Huesca 2016-2018. *Revista Clínica de Medicina de Familia*, 13(3), 198-202. <https://scielo.isciii.es/pdf/albacete/v13n3/1699-695X-albacete-13-03-198.pdf>
- Bolt, H., Suffel, A., Matthewman, J., Sandmann, F., Tomlinson, L., & Eggo, R. (2023). Seasonality of acute kidney injury phenotypes in England: An unsupervised machine learning classification study of electronic health records. *BMC Nephrology*, 24(1), 234. <https://doi.org/10.1186/s12882-023-03269-0>
- Cavagnaro Santa María, F. (2014). Resistencia antibiótica en la infección urinaria: la historia sin fin. *Boletín Médico del Hospital Infantil de México*, 71(6), 329-331. <https://doi.org/10.1016/j.bmhmx.2014.12.001>
- Chacaguasay, M., Reátegui, R., Valdiviezo-Díaz, P., & Chicaiza, J. (2024). Unsupervised techniques to identify patterns in gynecologic information. En G. Li, J. Filipe y Z. Xu (Eds.), *Communications in Computer and Information Science* (pp. 31-43). https://doi.org/10.1007/978-3-031-58956-0_3
- Elgoibar, B., Gangoiti, I., Garcia-Garcia, J. J., Hernandez-Bou, S., Gomez, B., Martinez, L., & Mintegi, S. (2020). Paediatric *Escherichia coli* bacteraemia presentations and high-risk factors in the emergency department. *Acta Paediatrica*, 110(3), 1032-1037. <https://doi.org/10.1111/apa.15549>
- Hadikurniawati W., Hartomo K. D., & Sembiring, I. (2023, 24 de noviembre). Spatial clustering of child malnutrition in central Java: A comparative analysis using k-means and DBSCAN. En *2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)* (pp. 242-247). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/icmeralda60125.2023.10458202>
- Hevia, P., Alarcón, C., Gonzalez, C., Nazal, V., & Rosati, M. P. (2020). Recomendaciones sobre diagnóstico, manejo y estudio de la infección del tracto urinario en pediatría. Rama de Nefrología de la Sociedad Chilena de Pediatría. Parte 1. *Revista Chilena de Pediatría*, 91(2), 281-288. <https://doi.org/10.32641/rchped.v91i2.1267>
- Hobensack, M., Zhao, Y., Scharp, D., Volodarskiy, A., Slotwiner, D., & Reading, M. (2023). Characterising symptom clusters in patients with atrial fibrillation undergoing catheter ablation. *Open Heart*, 10(2), e002385. <https://doi.org/10.1136/openhrt-2023-002385>
- Josephson, C. B., Gonzalez-Izquierdo, A., Engbers, J. D., Denaxas, S., Delgado-Garcia, G., Sajobi, T. T., Wang, M., Keezer, M. R., & Wiebe, S. (2023). Association of comorbid-socioeconomic clusters with mortality in late onset epilepsy derived through unsupervised machine learning. *Seizure: European Journal of Epilepsy*, 111, 58-67. <https://doi.org/10.1016/j.seizure.2023.07.016>

- Krueger, C., Alqurashi, W., Barrowman, N., Litwinska, M., & Le Saux, N. (2024). The long and the short of pediatric emergency department antibiotic prescribing: A retrospective observational study. *The American Journal of Emergency Medicine*, 75, 131-136. <https://doi.org/10.1016/j.ajem.2023.10.052>
- Leung, A. K. C., Wong, A. H. C., Leung, A. A. M., & Hon, K. L. (2019). Urinary tract infection in children. *Recent Patents on Inflammation & Allergy Drug Discovery*, 13(1), 2-18. <https://doi.org/10.2174/1872213X13666181228154940>
- Martins, C., Neves, B., Teixeira, A. S., Froes, M., Sarmiento, P., Machado, J., Magalhães, C. A., Silva, N. A., Silva, M. J., & Leite, F. (2024). Identifying subgroups in heart failure patients with multimorbidity by clustering and network analysis. *BMC Medical Informatics and Decision Making*, 24(1), artículo 95. <https://doi.org/10.1186/s12911-024-02497-0>
- Mohanty, S., Hassan, F. M., Lenke, L. G., Lewerenz, E., Passias, P. G., Klineberg, E. O., Lafage, V., Smith, J. S., Hamilton, D. K., Gum, J. L., Lafage, R., Mullin, J., Diebo, B., Buell, T. J., Kim, H. J., Kebaish, K., Eastlack, R., Daniels, A. H., Mundis, G., ... Bess, S. (2024). Machine learning clustering of adult spinal deformity patients identifies four prognostic phenotypes: A multicenter prospective cohort analysis with single surgeon external validation. *The Spine Journal*, 24(6), 1095-1108. <https://doi.org/10.1016/j.spinee.2024.02.010>
- Salamzade, R., McElheny, C. L., Manson, A. L., Earl, A. M., Shaikh, N., & Doi, Y. (2023). Genomic epidemiology and antibiotic susceptibility profiling of uropathogenic *Escherichia coli* among children in the United States. *mSphere*, 8(5), e00184-23. <https://doi.org/10.1128/msphere.00184-23>
- Shaikh, N., & Hoberman, A. (2022). *Urinary tract infections in children: Epidemiology and risk factors*. UpToDate. https://www.uptodate.com/contents/urinary-tract-infections-in-children-epidemiology-and-risk-factors?search=infection%20urinary%20children&source=search_result&selectedTitle=2~150&usage_type=default&display_rank=2#H2
- Suwono, B., Eckmanns, T., Kaspar, H., Merle, R., Zacher, B., Kollas, C., Weiser, A. A., Noll, I., Feig, M., & Tenhagen, B. (2021). Cluster analysis of resistance combinations in *Escherichia coli* from different human and animal populations in Germany 2014-2017. *PLoS ONE*, 16(1), e0244413. <https://doi.org/10.1371/journal.pone.0244413>
- Wazzan, A., Taconne, M., Le Rolle, V., Inngjerdigen, M., Hermann, K., Galli, E., Hernandez, A., Edvardsen, T., & Donal, E. (2024). Risk profiles for ventricular arrhythmias in hypertrophic cardiomyopathy through clustering analysis including left ventricular strain. *International Journal of Cardiology*, 409, 132167. <https://doi.org/10.1016/j.ijcard.2024.132167>

Whelan, S. O., Kyne, S., Dore, A., Glynn, M., Higgins, F., Hanahoe, B., Moriarty, F., Moylett, E., & Cormican, M. (2024). Paediatric *Escherichia coli* urinary tract infection: Susceptibility trends and clinical management. A retrospective analysis of a 10-year period. *Irish Journal of Medical Science*, 193, 1891-1900. <https://doi.org/10.1007/s11845-024-03670-0>