# Metodología de gestión de datos para identificar Síndrome de Intestino Irritable usando Aprendizaje de Máquina

Data Management Methodology to Identify Irritable Bowel Syndrome Using Machine Learning

Israel Cinta Ramírez[1], Arturo Yosimar Jaen Cuellar[1] Roberto Carlos Álvarez Martínez[1] Juan Pablo Amézquita Sánchez[1]
[1]Universidad Autónoma de Querétaro
icinta07@alumnos.uaq.mx[1]

## Resumen

El desarrollo de la inteligencia artificial nos permite analizar grandes cantidades de datos y realizar su clasificación, por lo cual se ha empezado a utilizar en distintos campos, incluso en el campo de la medicina. En dicho ámbito, la detección de algunas enfermedades presenta múltiples desafíos, ya que depende de factores como la experiencia del médico, la correcta identificación de indicadores biológicos, y la similitud con otras patologías. Estos desafíos contribuyen a que muchas enfermedades no sean diagnosticadas adecuadamente, como es el caso del Síndrome de Intestino Irritable (SII), una afección que impacta a miles de personas en todo el mundo y que carece de marcadores biológicos claros para su diagnóstico preciso. Este trabajo propone la aplicación de tres técnicas de Aprendizaje Máquina (AM) en bases de datos libres ya existentes de abundancias bacterianas del intestino en conjunto con algunos metadatos del paciente y 5 índices de diversidad biológica, para así diagnosticar el SII. El trabajo busca evitar las complicaciones asociadas con los métodos de diagnóstico tradicionales.

**Palabras clave**: Inteligencia Artificial, redes neuronales, diagnóstico, síndrome de intestino irritable, disbiosis, gestión de datos, aprendizaje máquina.

## Introducción

El Síndrome de Intestino Irritable (SII) es uno de los trastornos intestinales más comunes, afectando al 4.1% de la población mundial. Se caracteriza por ocasionar dolor abdominal, meteorismo con distensión abdominal y alteraciones en las evacuaciones intestinales. Este trastorno tiene una prevalencia del 10% al 25% en Estados Unidos de América, entre un 17% y un 21% en América del Sur, del 7% al 9% en el sur de Asia, y del 5.6% en Medio Oriente y África. Dentro de la población afectada por el SII, hay personas de todas las edades, pero este síndrome tiene una mayor prevalencia en personas de entre 32 y 38 años [1]. Debido a que afecta a una gran parte de la población, y en especial a personas en edades donde se concentra la mayor parte de la fuerza laboral, es la segunda causa de ausentismo laboral después de la gripe. En Estados Unidos, provoca un gasto de entre 2.4 y 3.5 millones en consultas médicas al año, 2.2 millones de dólares en prescripciones, y un gasto de 8 mil millones de dólares en costos directos [2]. Estos gastos se deben en gran parte a que es un síndrome cuya causa no se explica por alteraciones morfológicas, metabólicas o neurológicas demostrables mediante las técnicas de diagnóstico habituales. La falta de utilización de indicadores biológicos hace que el diagnóstico del SII sea difícil y, en algunos casos, se descarten otras enfermedades [3]. Por estas razones, el gasto económico para tratar el SII es considerable, ya que solo se suelen tratar los síntomas y no la enfermedad debido a la inexistencia de un tratamiento y la falta de información sobre esta [4]. Por estas problemáticas, es importante explorar en el desarrollo de metodologías que permitan diagnosticar adecuadamente este tipo de enfermedades, de una forma automática, basada en datos, indicadores, métricas, y características que sean insensibles a la interpretación como los métodos tradicionales hacen.

Debido a las dificultades de aplicar la información que se conoce sobre el SII para desarrollar métodos para un diagnóstico, generalmente se utilizan los criterios de Roma, los cuales son un listado de síntomas que deben persistir durante al menos 3 meses [5]. De manera más profunda, se ha estudiado el SII para encontrar diferencias entre la microbiota intestinal de personas saludables y personas con el SII, donde se enlistan las familias de bacterias y las diferencias que tienen entre los dos grupos, además de aplicar análisis de componentes principales (PCA, por sus siglas en inglés) para diferenciarlos [6]. En un trabajo similar, se comparan las abundancias de bacterias de pacientes con SII que presentan distintos síntomas y pacientes

sanos, utilizando indicadores estadísticos como la varianza y la media de los datos. Además, se relacionan las abundancias de ciertas familias de bacterias con los síntomas del SII [7]. Teniendo en cuenta todo esto, se puede observar que los métodos de diagnóstico del SII siguen basándose en los cuestionarios de los criterios de Roma, aunque haya estudios que señalan cómo la disbiosis es un parámetro con el cual se puede distinguir un paciente sano de uno con SII. Sin embargo, estos estudios no suelen utilizarse para el diagnóstico, posiblemente porque el análisis de datos puede ser demasiado extenso y complejo para un médico, por lo que un método automatizado podría ser de gran ayuda.

Los métodos de aprendizaje máquina (ML, por sus siglas en inglés) son programas a los que se les "enseña" a reconocer patrones, y una vez aprendidos, estos programas se pueden utilizar para analizar grandes cantidades de datos. Estos métodos se han utilizado para prevenir la enfermedad periodontal, donde se emplearon en mayor porcentaje los algoritmos de Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) y Aprendizaje Profundo (DL, por sus siglas en inglés). Aunque se menciona que los algoritmos y métodos varían dependiendo de la aplicación, la mayoría tienden a tener buenos resultados de precisión, superiores al 70% [8]. En un trabajo más enfocado en el SII, se propusieron 108 variables, entre las que se incluyen hospitalizaciones, comorbilidades y pruebas de laboratorio, para entrenar una IA que identifique la enfermedad inflamatoria intestinal, utilizando los algoritmos de regresión Ridge, regresión LASSO (Least Absolute Shrinkage and Selection Operator), SVM, Bosque Aleatorio y Redes Neuronales Artificiales (ANN, por sus siglas en inglés). Los métodos de Bosque Aleatorio y de regresión LASSO obtuvieron los mejores resultados, con una precisión de entre 70% y 92% [9]. De manera similar a la metodología propuesta, se ha utilizado IA para distinguir entre un paciente saludable y uno con cáncer de colon al analizar su microbiota, comparando la abundancia de distintas especies de bacterias y elaborando un listado de 20 especies que disminuyeron en abundancia y 20 que aumentaron en los pacientes con cáncer de colon [10]. Como se aprecia en los trabajos anteriores, utilizar métodos de ML para diagnosticar patologías puede arrojar resultados con buena precisión y de manera rápida, lo que puede ser de gran ayuda para evitar problemas como el manejo de grandes cantidades de datos, la falta de experiencia o el error humano, como podría ser el caso en el diagnóstico del SII, donde los métodos basados en datos de cuestionarios han mostrado buenos resultados.

En el presente trabajo, se propone una metodología basada en gestión de datos y ML para detectar de forma sencilla, objetiva, rápida y automática el SII usando indicadores basados en la microbiota intestinal. La metodología se aplicará en bancos de datos, de acceso libre, que contienen abundancias de bacterias en el intestino de personas con SII. Se emplearán tres métodos de ML: SVM, K-Vecinos más Cercanos (K-Nearest Neighbours o KNN) y Redes Neuronales Artificiales (Artificial Neural Networks o ANN). Estas técnicas se evaluarán utilizando métricas de desempeño como precisión, exactitud, sensibilidad y F1. El objetivo de la aplicación de estos métodos es diferenciar a un paciente con SII de uno sano, obteniendo mejores resultados que con los cuestionarios de los criterios de Roma y, lo más importante, basándose en datos cuantitativos. En el proceso de diagnóstico, se obtienen las siguientes características de la base de datos para facilitar su análisis: los indicadores estadísticos de los filos de las abundancias de bacterias de la base de datos y cinco índices de diversidad biológica más comunes. Los índices de diversidad biológica no se suelen utilizar para propósitos de diagnóstico, pero pueden servir como indicadores clave en este tipo de análisis, ya que condensan la información de las abundancias de bacterias. Por ello, el uso de ML es crucial, ya que esta herramientas permite procesar grandes cantidades de datos en poco tiempo, además de reducir los errores de origen humano.

## Fundamentación teórica

Las enfermedades tienden a mostrar una disbiosis en el intestino, en especial las enfermedades gastrointestinales, pero analizar los datos de disbiosis es demasiado complicado para una persona o grupo de personas, sin mencionar el error humano, por lo que se utilizan los métodos de ML para facilitar el análisis de la base de datos e incluso se obtienen los índices de diversidad biológica de las abundancias de bacterias para que sean otras características de análisis.

### Disbiosis

La disbiosis se puede describir como una alteración compositiva y funcional de la microbiota intestinal causada por un factor externo o del individuo. Aunque la microbiota tiende a adaptarse a los cambios que se presentan, como pueden ser los picos de estrés, la mala alimentación, el uso de medicamentos o el sufrir de

alguna enfermedad, cuando la afectación ha desaparecido y la composición microbiana no ha regresado a su estado normal y persiste crónicamente tiende a tener consecuencias perjudiciales para el huésped, [11].

## Descripción del SII

El SII es un síndrome que se caracteriza por la presencia de dolor abdominal recurrente asociado a alteraciones del ritmo deposicional, ya sea en forma de estreñimiento, de diarrea, o de ambas, la hinchazón y la distención abdominal también son comunes. Debido a que no tiene indicadores biológicos además de los síntomas se le suele confundir con otras enfermedades gastrointestinales, razón por la cual afecta a muchas personas de manera crónica.

## Banco de datos de bacterias

Una base de datos de bacterias contiene muestras de la microbiota, principalmente del intestino, de cientos de personas con diferentes hábitos, orígenes y patologías. Existen muchas bases de datos de bacterias de acceso libre, disponibles para su uso en investigaciones y desarrollos, siempre y cuando se otorguen los respectivos agradecimientos.

La base de datos utilizada en este trabajo se obtuvo de la página web https://www.ebi.ac.uk/. En esta página se pueden encontrar diversos estudios, artículos y trabajos junto con los datos que utilizaron. Las bases de datos están disponibles en formatos ".sra" y ".lite.1", los cuales se pueden convertir a formato ".csv". Una vez convertidas, las bases de datos contienen información no personal de los pacientes, como el tipo de dieta, sintomatología, edad, sexo, antecedentes de uso de tabaco, etc., las abundancias de bacterias, y los nombres de las especies bacterianas.

La base de datos utilizada en este trabajo contiene datos obtenidos del análisis de la microbiota fecal de 26 personas, de las cuales 20 tienen un diagnóstico de Síndrome de Intestino Irritable (SII) y 6 son personas sanas que sirven como grupo de control.

## Índices de diversidad

Se les conoce como índices de diversidad a la representación matemática de la diversidad biológica de una zona determinada, basándose en la cantidad de especies y el número de individuos de cada especie. Cada índice tiene una fórmula distinta debido a que cada autor basó su fórmula en distintos criterios, por lo cual se suelen utilizar varias fórmulas para que se puedan contrastar resultados con otros estudios.

Los índices de diversidad más utilizados son: La riqueza de especies, el índice de Shannon, Exponencial de índice de Shannon, Índice de Gini-Simpson e inverso de Gini-Simpson, estos se observan en las ecuaciones (1) a (5) respectivamente.

$$D_{rich} = S = \sum p_i^0 \tag{1}$$

$$H_{Shannon} = -\sum p_i * log_b(p_i) \tag{2}$$

$$D_{\exp Shannon} = b^{H_{Shannon}} \tag{3}$$

$$H_{Gini-Simpson} = 1 - \sum p_i^2 \tag{4}$$

$$D_{inv\ Gini-Simpson} = \frac{1}{1 - H_{Gini-Simpson}} = \frac{1}{\sum p_i^2} \tag{5}$$

Donde $pi$ es la abundancia relativa de la especie $i$, es decir la abundancia de la especie $i$ dividida entre la suma de las abundancias de las $S$ especies de la comunidad analizada, y $b$ es la base del logaritmo con el que se está trabajando [12].

## Indicadores estadísticos

Para el análisis de datos es necesario obtener la mayor cantidad de propiedades, de manera que se puedan usar como comparación. Es por eso que se suelen utilizar ciertos indicadores que permiten al sistema de aprendizaje distinguir entre los conjuntos de datos [13]. En las ecuaciones (6) a (8) se muestran los indicadores estadísticos utilizados: Media, varianza y desviación estándar respectivamente.

$$\frac{1}{N}\sum_{i=1}^{N} x_i \tag{6}$$

$$\frac{1}{N} * \sum_{i=1}^{N} (x_i + \bar{x})^2 \tag{7}$$

$$\sqrt{\frac{1}{N} * \sum_{i=1}^{N} (x_i + \bar{x})^2} \tag{8}$$

Donde *N* es el número de datos, *x* es el dato, *i* es el número de dato actual y $\bar{x}$ es la media de los datos.

## Clasificadores basados en IA

Cuando se habla de IA se suele hablar también del ML, que es un método basado en la supervisión de la aplicación de un algoritmo de aprendizaje con una arquitectura específica para la clasificación de datos.

Entre los algoritmos de aprendizaje están los Árboles de Decisión (en inglés Decision Trees), SVM, KNN y ANN [10].

## Máquinas de Soporte Vectorial

Las SVM se definen como sistemas que usan el espacio hipotético de funciones lineales en un espacio dimensional superior para maximizar el margen entre el hiperplano y los dos puntos más cercanos de las categorías que se quieren separar. Siendo el hiperplano un subespacio cuya dimensión es una menor al espacio que componen los datos.

Los conjuntos de datos deben situarse a un lado u otro de los márgenes del hiperplano, nunca entre ambos como se observa en la Fig. 1.

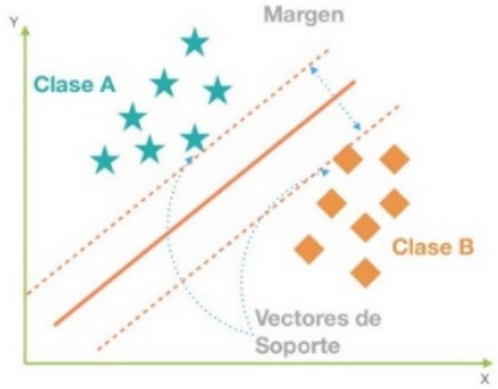


*Figura 1 Descripción gráfica de SVM. Fuente: Gonzalez L. 2019, aprendeia.com*

## K Vecinos Más Cercanos

Este algoritmo es un método que se basa en la asunción de que una instancia tiene las mismas características que los k elementos de instancias que están más cerca de esta. Para medir la distancia entre las instancias se suele utilizar la distancia euclídea, la cual se muestra en la ecuación 7 [10]. En la Fig. 2 se aprecia gráficamente el método.

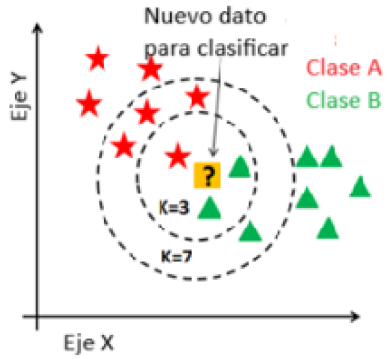


*Figura 2 Representación gráfica del método de KNN. Fuente: Gomez W. J. 2022, rstudio-pubs-static.s3.amazonaws.com*

## Redes Neuronales Artificiales – Perceptrón Multicapa

Las ANN son modelos de aprendizaje automático que recrean modelos simples del cerebro humano para resolver problemas complejos. El Perceptrón Multicapa (MLP, por sus siglas en inglés) es una red neuronal que está compuesta por varias capas de neuronas, donde las neuronas de la primera capa son la fila de las neuronas de entrada donde llegan los valores de entrada de la red, la última capa son la o las neuronas de salida que proporcionan él o los resultados y las capas ocultas son las capas intermedias que no están conectadas con ninguna entrada o ninguna salida [14]. En la Fig. 3 se muestra una neurona artificial siendo X1 y X2 los datos de entrada, W11, W12 y W13 los pesos a calcular por la IA y hjp el valor resultante que proporciona la neurona.

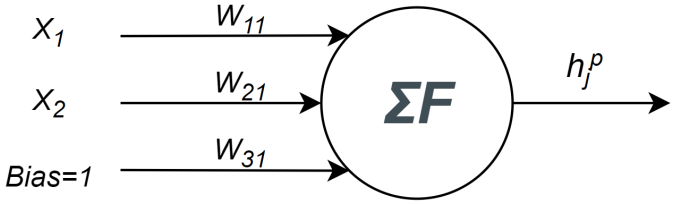


*Figura 3 Representación gráfica de una neurona artificial (Autoría propia)*

## Análisis de Componentes Principales

El Análisis de Componentes Principales (en inglés Principal Components Analysis o PCA) es una técnica que se suele aplicar en el análisis exploratorio de datos. Una de sus principales aplicaciones es la de la reducción de dimensionalidad perdiendo la menor cantidad de información posible, lo cual nos permite la visualización de datos con una gran cantidad de variables [13].

## Métricas de desempeño

Es importante tener métricas de rendimiento para discriminar diferentes algoritmos de ML. Dichas métricas de desempeño se consiguen aplicando una matriz de confusión para variables dicotómicas [13].

“Accuracy” (en inglés) o exactitud nos permite saber que tan efectivo es nuestro algoritmo en acercarse al valor real de la clasificación.

“Recall” (en inglés) o sensibilidad que nos da a conocer la proporción de casos positivos que fueron correctamente clasificados.

“Precision” (en inglés) o precisión nos permite saber que tan cerca están los valores conseguidos entre sí, o también la repetitividad que tiene nuestro algoritmo tiene para dar de nuevo el mismo resultado con el mismo dato.

“F1” es una medida que combina “accuracy” y “recall” por lo que nos dice que tanto se acerca nuestro algoritmo a los casos reales y positivos.

## Metodología

A continuación, en la Fig. 4, se muestra el diagrama general de la metodología propuesta para identificar personas con SII usando gestión de datos y ML:

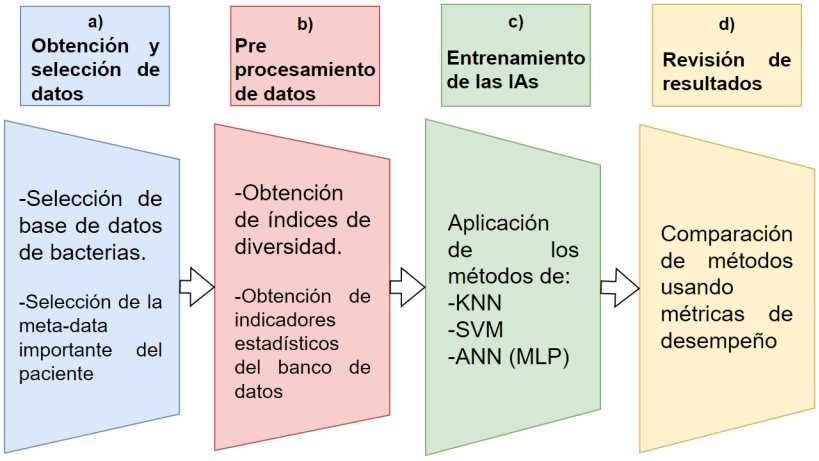


*Figura 4 Diagrama de la metodología (Autoría propia)*

Los pasos generales son cuatro y se retoman más a fondo a continuación.

a) Obtención de las bases de datos y selección de los datos:
La base de datos elegida se compone de 26 casos, de las cuales 20 personas tienen SII y 6 personas sirven como grupo de control. De entre los metadatos se seleccionan las variables de dieta y si son fumadores, ya que el resto de los datos hacían más difícil el diagnóstico, las abundancias de bacterias y los nombres científicos de estas, que incluyen hasta 645 especies. Los datos se manejan en formato .csv.

b) Obtención de índices de diversidad:
Se calculan cinco índices de diversidad, utilizando las ecuaciones (1) a (5), a partir de las abundancias de bacterias, los cuales condensan la información de toda la base de datos. Además, se obtienen los valores de la media, varianza y desviación estándar de las abundancias de cada filo de bacterias, utilizando las ecuaciones (6) a (8), lo que permite condensar la información de la base de datos, pero aun así enfocarse en información específica. Se usaron solo estos indicadores estadísticos ya que el utilizar más hacían difícil el diagnóstico o no hacían ninguna diferencia en el análisis. El uso de estos índices y valores estadísticos reduce el análisis de las 645 especies de bacterias a solo 21 variables.

c) Entrenamiento de la IA:
Se utilizan los índices de diversidad y los promedios de abundancias para entrenar los algoritmos basados en tres métodos de Machine Learning (ML). Se aplica el método de Análisis de Componentes Principales (PCA) para reducir los datos a dos dimensiones y así implementar K-Nearest Neighbors (KNN), ya que este método presenta dificultades en problemas con alta dimensionalidad cuando se utiliza una distancia euclidiana. Para los métodos de Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales Multicapa (ANN-MLP), se emplean todas las variables. Los datos se dividen en dos conjuntos: 14 casos para entrenamiento y 12 casos como banco de pruebas. Esta distribución se eligió porque el número total de datos no era lo suficientemente grande como para destinar el 80% al entrenamiento, además de que podrían surgir

problemas de sobreajuste con una muestra tan pequeña. Por ello, se optó por un número ligeramente superior al 50% de los datos para el entrenamiento.

d) Revisión de resultados:
Se prueba la capacidad de los métodos utilizando el banco de pruebas, y se obtienen las métricas de desempeño correspondientes.

## Resultados y discusión

Al aplicar PCA sobre la base de datos para disminuir el número de variables se obtuvo la gráfica que se observa en la Fig. 5 es una proyección de dos dimensiones de la base de datos, representados como dos características.

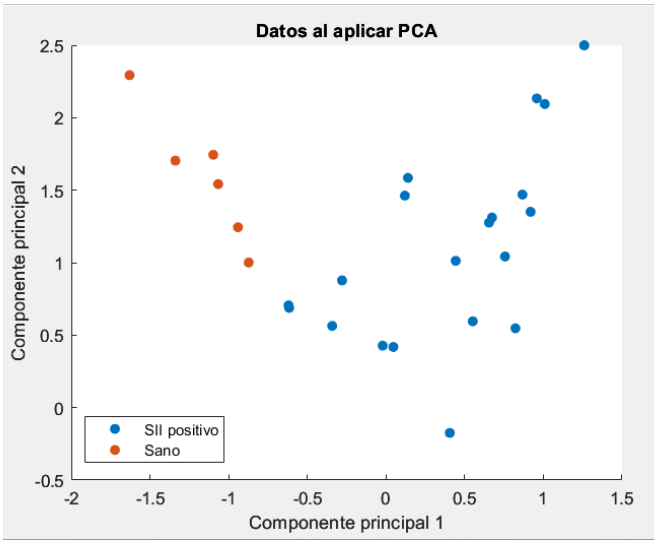


*Figura 5 Base de datos resultante al pasar por PCA (Autoría propia)*

Al usar los datos obtenidos con PCA para entrenar y aplicar el método de KNN se obtuvieron los resultados observados en la Fig. 6.

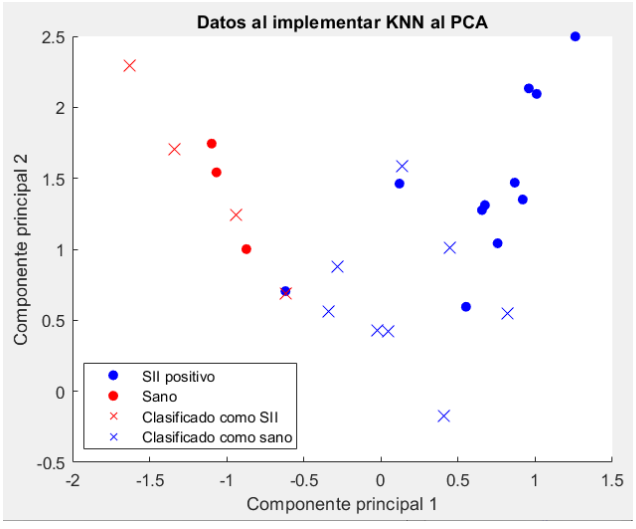


*Figura 6 Gráfico de clasificación del método de KNN (Autoría propia)*

Los resultados del KNN se aprecian en la tabla 1.
***Tabla 1*** *Resultados de clasificación del KNN*

| | PREDICCIÓN | | |
|---|---|---|---|
| **VERDADERO** | | Positivo | Negativo |
| | Positivo | 8 | 1 |
| | Negativo | 0 | 3 |

Por lo tanto, las métricas de desempeño obtenidas con los resultados del KNN se observan en la tabla 2.
***Tabla 2*** *Métricas de desempeño del KNN*

| | |
|---|---|
| Exactitud | 91.66% |
| Precisión | 100% |
| Sensibilidad | 88.89% |
| F1 | 94.12% |

En el caso de la SVM se obtuvieron los resultados que se aprecian en la tabla 3.
***Tabla 3*** *Resultados de clasificación del SVM*

| | Predicción | | |
|---|---|---|---|
| Verdadero | | Positivo | Negativo |
| | Positivo | 8 | 1 |
| | Negativo | 0 | 3 |

Por lo que las métricas de desempeño del SVM se observan en la tabla 4.
***Tabla 4*** *Métricas de desempeño del SVM*

| | |
|---|---|
| Exactitud | 91.67% |
| Precisión | 100% |
| Sensibilidad | 88.89% |
| F1 | 94.12% |

En el caso de la ANN-MLP se obtuvieron los resultados que se observan en la tabla 5.
***Tabla 5*** *Resultados de clasificación del ANN-MLP*

| | Predicción | | |
|---|---|---|---|
| Verdadero | | Positivo | Negativo |
| | Positivo | 7 | 0 |
| | Negativo | 2 | 3 |

Las métricas de desempeño del ANN-MLP se pueden observar en la tabla 6.
***Tabla 6*** *Métricas de desempeño del ANN-MLP*

| | |
|---|---|
| Exactitud | 83.33% |
| Precisión | 77.78% |
| Sensibilidad | 100% |
| F1 | 87.5% |

La comparación de los métodos se muestra en la Fig. 7. Donde se aprecia mejor como los métodos de KNN y SVM tienen las mismas métricas de desempeño y ambos superan en exactitud, precisión y F1 al método de ANN. Siendo la sensibilidad la única métrica en la que ANN es superior a los otros dos métodos.

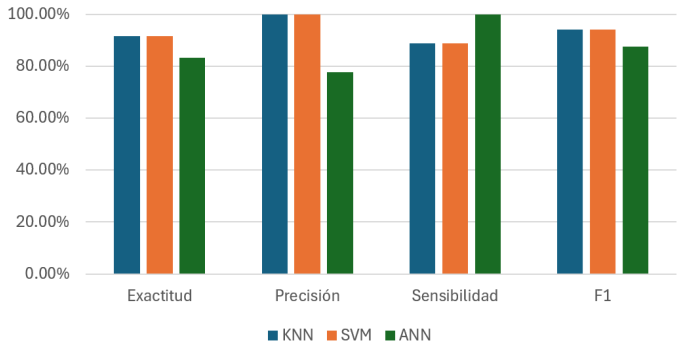


*Figura 7 Métricas de desempeño (Autoría propia)*

## Conclusiones

Los métodos de ML tuvieron buenas métricas de desempeño en general obteniendo valores superiores al 80%, pero los métodos de KNN y SVM tuvieron los mejores resultados a la hora de diagnosticar el SII con la base de datos utilizada.

La similitud de resultados del KNN y el SVM podría deberse a la cercanía que tienen algunos datos lo cual cause confusión a la hora de catalogar los datos más cercanos a la “frontera que los divide”.

El que la ANN-MLP no haya dado buenos resultados en comparación al KNN y al SVM tal vez se deba a la falta de datos para su entrenamiento. Mientras que el SVM puede distinguir las fronteras de los 2 grupos con pocos datos y el KNN puede hacer una clasificación lógica gracias a la cercanía de los datos, aunque sean pocos.

Entre los metadatos que proporciona esta base de datos en particular la información de edad, sexo y el tipo de heces no proporcionaban información útil para el diagnóstico de los casos por lo que se utilizaron solo los datos de la dieta y los antecedentes de tabaquismo. En el caso de los indicadores estadísticos, los datos de media fueron los que proporcionaban una mayor capacidad de discriminación a la hora de utilizar los métodos de ML, los indicadores de mediana y varianza tenían una capacidad menor de discriminación, pero el resto de los indicadores como la desviación estándar y la moda reducían la capacidad de discriminación de los métodos de ML.

Utilizando la metodología planteada se lograron obtener buenos resultados en el diagnóstico del SII utilizando los datos de disbiosis. Siendo los mejores algoritmos los de KNN y SVM. Por lo que podría utilizarse a mayor escala para detectar la enfermedad y se puede esperar que estos algoritmos vuelvan a tener buenas métricas.

Como retrospectiva se podría utilizar una base de datos más amplia para futuros análisis.

## Agradecimientos

Se agradece al Consejo Nacional de Humanidades, Ciencias y Tecnologías por los recursos otorgados.

Se agradece a EMBL-EBI por compartir la base de datos.

## Referencias

[1] Domingo, J. J. S. Síndrome del intestino irritable. *Medicina Clínica (English Edition)*. 158(2), 76-81. 2021.

[2] Shaikh, S.D.; Sun, N.; Canakis, A.; Park, W.Y.; Weber, H.C. Irritable bowel syndrome and the gut microbiome: A comprehensive review. *Journal of Clinical Medicine*. 12(7). 2023.

[3] Otero W. y Gómez M. Síndrome de Intestino Irritable: diagnóstico y tratamiento farmacológica Revisión concisa. *Revista de gastroenterología del Perú*. 25. 2005.

[4] Ghaffari P., Shoaie S., Nielsen L. Irritable bowel syndrome and microbiome; Switching from conventional diagnosis and therapies to personalized intervetions. *J Transl Med.* 20, 173. 2022.

[5] Mearin, F., Ciriza, C., Mínguez, M., Rey, E., Mascort, J. J., Peña, E., ... & Júdez, J. (2016). Guía de Práctica Clínica: Síndrome del intestino irritable con estreñimiento y estreñimiento funcional en adultos. *Revista Española de Enfermedades Digestivas*, *108*(6), 332-363.

[6] Shukla, R., Ghoshal, U., Dhole, T. N., & Ghoshal, U. C. (2015). Fecal microbiota in patients with irritable bowel syndrome compared with healthy controls using real-time polymerase chain reaction: an evidence of dysbiosis. *Digestive diseases and sciences*, *60*, 2953-2962.

[7] Pozuelo, M., Panda, S., Santiago, A., Mendez, S., Accarino, A., Santos, J., ... & Manichanh, C. (2015). Reduction of butyrate-and methane-producing microorganisms in patients with Irritable Bowel Syndrome. *Scientific reports*, *5*(1), 12693.

[8] González Lozada, L., & Pinzón Murillo, J. A. (2024). Herramientas De Inteligencia Artificial Para La Prevención De Enfermedad Gingival Y Periodontal: Revisión De Alcance.

[9] Zand A., Stokes Z., Sharma A., Van Deen W. y Hommes D. Artificial Intelligence for Inflammatory Bowel Diseases (IBD); Accurately Predicting Adverse Outcomes Using Machine Learning. *Dig Dis Sci.* 67, 4874–4885. 2022.

[10] Patón González, V. Uso de inteligencia artificial para prevenir la formación de fístulas colónicas y su relación con la microbiota intestinal [Trabajo Fin de Grado, E.T.S. de Ingeniería Agronómica, Alimentaria y de Biosistemas (UPM)]. 2021.

[11] Moreno X. Disbiosis en la microbiota intestinal. *Revista GEN*. 76(1), 17-23. 2022.

[12] González J. Midiendo la diversidad biológica: más allá del índice de Shannon. *Acta zoológica lilloana*, 3-14. 2012.

[13] Roweis, S. (1997). EM algorithms for PCA and SPCA. *Advances in neural information processing systems*, *10*.

[14] Huerta-Rosales, J. R., Granados-Lieberman, D., Garcia-Perez, A., Camarena-Martinez, D., Amezquita-Sanchez, J. P., & Valtierra-Rodriguez, M. (2021). Short-circuited turn fault diagnosis in transformers by using vibration signals, statistical time features, and support vector machines on FPGA. *Sensors*, *21*(11), 3598.

[15] Matich D. Redes Neuronales: Conceptos Básicos y Aplicaciones. *Acta zoológica lilloana*, 3-14. 2001.