



## Diagnóstico de infertilidad utilizando técnicas de aprendizaje profundo: un enfoque prometedor en medicina reproductiva

### Diagnosis of infertility using deep learning techniques: a promising approach in reproductive medicine

**Mario Andrés Basilio López**

Universidad del Istmo, campus Tehuantepec, Santo Domingo Tehuantepec Oaxaca, México  
andres.basilio@outlook.es

**Guillermo Eduardo Reyes Rodríguez**

Universidad del Istmo, campus Tehuantepec, Santo Domingo Tehuantepec Oaxaca, México  
eduar1299@hotmail.com

**José Ricardo Salvador Nolasco**

Universidad del Istmo, campus Tehuantepec, Santo Domingo Tehuantepec Oaxaca, México  
jose\_rivel@hotmail.com

**Sergio Juárez Vázquez**

Universidad del Istmo, campus Tehuantepec, Santo Domingo Tehuantepec Oaxaca, México  
sjuarez@sandunga.unistmo.edu.mx  
ORCID: 0000-0002-2080-4861

doi: <https://doi.org/10.36825/RITI.11.23.006>

Recibido: Febrero 20, 2023

Aceptado: Mayo 03, 2023

**Resumen:** La disminución de la tasa de fertilidad mundial es una preocupación, ya que existe la posibilidad de que la disminución de la tasa de fertilidad mundial pueda llevar a la extinción humana. La infertilidad en hombres es un tema de interés para la salud pública, debido a la disminución de la cantidad y concentración de espermatozoides, causada por hábitos de vida, enfermedades y accidentes. La identificación de factores de riesgo en la infertilidad en hombres no es común en la medicina, pero se están desarrollando técnicas de inteligencia artificial para ayudar en la identificación de estos factores de riesgo. Este artículo se enfoca en el análisis de datos mediante minería de datos, usando la base de datos *Fertility*, que contiene información de 100 voluntarios. La técnica escogida para el análisis de datos fueron los árboles de decisión en MATLAB. La base de datos original tenía 9 atributos y se redujo a 5 para el modelo de clasificación. En la segunda etapa, el entrenamiento y evaluación cruzada del modelo de clasificación obtuvo un 83.3% de precisión y un tiempo de entrenamiento de 1.8774 segundos. La última etapa, una prueba con el 10% de las muestras, obtuvo un 80% de aciertos. El modelo tuvo una Tasa de Verdaderos Positivos del 94.9% para la clase N y una Tasa de Falsos Negativos del 100% para la clase O.

**Palabras clave:** *Infertilidad, Árboles de Decisión, Análisis de Datos, Minería de Datos, Matriz de Confusión.*

**Abstract:** The decrease in the worldwide fertility rate is a concern due to the potential risk of an aging population and demographic imbalances. Male infertility is a matter of interest to public health officials and researchers, as there has been a decrease in sperm count and motility caused by lifestyle habits, diseases, and accidents. The identification of risk factors in male infertility is not common in medicine, however, researchers are developing artificial intelligence techniques to identify these risk factors. This article focuses on data analysis through data mining, using the Fertility database, which contains information from 100 volunteers. The technique chosen for data analysis was decision trees, implemented in MATLAB. The original database had 9 attributes and was reduced to 5 for the classification model. During the second stage, cross-training and evaluation produced a model accuracy of 83.3% and a training time of 1.8774 seconds. In the last stage, a test with 10% of the samples obtained 80% accuracy. The model produced a True Positive Rate of 94.9% for class N and a False Negative Rate of 100% for class O.

**Keywords:** Infertility, Decision Trees, Data Analysis, Data Mining, Confusion Matrix.

## 1. Introducción

La civilización está en riesgo debido a la infertilidad y la esterilidad, según datos estadísticos [1], la tasa de fertilidad mundial llegó a su punto más alto de 5.3% en los 60's y a partir de ahí se ha notado un decremento hasta el grado de tener una tasa de fertilidad de un 2.3% para el año 2020.

La infertilidad y la esterilidad son términos que se relacionan con la capacidad de una persona para concebir, pero ciertamente tienen diferencias importantes, la primera se refiere a la incapacidad permanente de concebir, mientras que la segunda hace referencia a la incapacidad temporal u ocasional de concebir, según [2] desde hace años existe un precipitado descenso del número de espermatozoides, preocupando a la humanidad de estar en el camino hacia la extinción.

Actualmente, la fertilidad en los hombres es un tema de interés dentro de la sociedad y se ve con ojos de preocupación, pues según [3] *“Los hombres van a tener problemas para ser fértiles en apenas una década, de seguir este ritmo: la cantidad de espermatozoides por eyaculación no ha dejado de bajar desde hace casi un siglo. Su concentración también ha bajado a menos de la mitad que hace 50 años, acercándose al umbral de la infertilidad. Y el ritmo del descenso se ha acelerado en lo que va de siglo, doblándose”*.

Es importante mencionar que la infertilidad puede ser causada por una variedad de factores: metabólicos, ambientales, hábitos alimenticios, sedentarismo, alcohol, tabaco, enfermedades y accidentes sufridos por las personas [4], por lo que el tratamiento dependerá de la causa subyacente. Es por ello que detectar los factores más influyentes en la fertilidad de una persona puede ayudar a identificar los patrones que derivan la baja de espermatozoides en la población y poder contrarrestar los efectos negativos.

Actualmente, la identificación de factores que contribuyen a la infertilidad masculina a través de técnicas computacionales no es una práctica comúnmente utilizada en la medicina. Sin embargo, se están desarrollando investigaciones que implementan técnicas de inteligencia artificial para ayudar en la identificación de factores de riesgo de la infertilidad masculina.

Entre las técnicas utilizadas para predecir la infertilidad se encuentran: el uso de redes neuronales, las cuales utilizan algoritmos [5] para analizar los datos de los pacientes, el análisis de imágenes de ecografías testiculares mediante técnicas de aprendizaje automático [6], y el análisis de grandes conjuntos de datos de pacientes con infertilidad, mediante técnicas de minería de datos [7, 8, 9], para identificar patrones y factores de riesgo. Para el presente estudio, se ha elegido la minería de datos, una técnica que permite descubrir relaciones, patrones y tendencias significativas al examinar grandes cantidades de datos [10]. El proceso de descubrimiento se realiza en varias etapas, que incluyen: la comprensión del dominio, preparación del conjunto de datos, descubrimiento de patrones, el análisis de patrones descubiertos y la utilización de resultados.

El objetivo de este estudio es evaluar los beneficios del uso de los árboles de decisión como algoritmo para estudiar los resultados de un diagnóstico de fertilidad, incluyendo el tiempo de ejecución, el número de casos clasificados correctamente y el porcentaje de error. De esta manera, se busca determinar si los árboles de decisión son más efectivos que otros métodos similares, al analizar los mismos casos de la base de datos *Fertility*, obtenida del repositorio UCI [11], Además, se espera que el empleo de la herramienta MATLAB permita una reducción en el número de factores necesarios para lograr un diagnóstico preciso, ya que como se menciona en [12] los estudios

indican que, en la mayoría de los grupos de investigación, es preferible trabajar con todas las variables, sin embargo, en ciertos casos, la reducción de variables puede resultar en pronósticos significativos.

## 2. Estado del arte

La minería de datos ha demostrado ser una técnica efectiva para el diagnóstico de la fertilidad en diversos estudios. Una de las estrategias más utilizadas es aplicar técnicas de aprendizaje automático para analizar datos de pacientes, como niveles hormonales, características físicas y antecedentes médicos, y así predecir la probabilidad de embarazo. Además, algunos estudios han utilizado técnicas de minería de datos para analizar imágenes de diagnóstico, como ecografías, con el fin de evaluar la calidad del óvulo y del espermatozoide.

En el momento de esta publicación, se identificaron artículos similares que trabajan con algoritmos de clasificación y conjuntos de datos relacionados con el estudio de los factores que afectan a la fertilidad de las personas. Aunque se han utilizado en varios estudios para predecir la fertilidad, aún es necesario realizar una gran cantidad de investigación para mejorar su precisión y validar sus resultados en diferentes grupos de población.

En [13] se mide el desempeño de diferentes algoritmos de clasificación aplicados a la base de datos de *Fertility* para identificar el mejor algoritmo. Se aplicaron dos tipos de *test* diferentes (*Crossvalidation* y *Use training set*) a cada algoritmo implementado. Además, se realizó una comparación en cuanto a la clasificación y tiempo para medir el desempeño de cada algoritmo, concluyendo que el algoritmo con mejor desempeño fue el *KStar*, pues tuvo un mayor número de instancias bien clasificadas y el menor tiempo de procesamiento en ambos tipos de pruebas.

En [14], se buscó encontrar el algoritmo con mayor efectividad empleando los algoritmos *KStar*, *Random Forest*, Regresión Lógica Bayesiana y Nnge para la clasificación a la base de datos *Fertility* extraída del repositorio UCI. Se emplearon estadísticos de kappa para analizar la concordancia y la reproducibilidad a cada algoritmo. Los resultados mostraron que tres de los cuatro algoritmos empleados obtuvieron una clasificación casi perfecta de las instancias. Sin embargo, el método con mayor velocidad de clasificación fue el *KStar*, con un tiempo de procesamiento de 0 s.

Por último, en [4] se busca predecir las variables más importantes que afectan la fertilidad, empleando los algoritmos *Clusters EM*, *SimpleKMeans* y *J48* para analizar los registros de la base de datos *Fertility* extraída del repositorio UCI, empleando técnicas de minería de datos. Se concluyó que las técnicas de árboles de decisión y los clústeres permiten identificar las variables más influyentes, dando como resultado las siguientes variables: edad, accidente, fiebre, cirugía y alcohol, que determinan las principales causas de infertilidad en las personas mayores de 23 años.

Los artículos [13, 14, 4] utilizaron el software WEKA para la realización de sus pruebas. Sin embargo, para la elaboración de las pruebas de este artículo, se utilizará el software de MATLAB.

## 3. Materiales y métodos

Con el propósito de analizar las características que influyen en un diagnóstico de fertilidad y determinar los factores más importantes, se utilizó la base de datos *Fertility* obtenida del repositorio UCI, la cual contiene el registro de 100 voluntarios, los cuales proporcionaron una muestra de semen analizada según los criterios de la OMS (Organización Mundial de la Salud).

Dicha base de datos está compuesta por un total de 8 características de entrada para el análisis y un atributo de salida que determina la clasificación. Los valores y explicación pueden verse en la Tabla 1.

**Tabla 1.** Atributos de la base de datos *Fertility*.

Características	Opciones	Valor
Estación en la que se realizó el análisis.	1) invierno 2) primavera 3) verano 4) otoño	-1 -0.33 0.33 1
Edad en el momento del análisis	18-36	0 (si) 1 (no)

Enfermedades infantiles (es decir, varicela, sarampión, paperas, poliomielitis)	1) sí 2) no.	0 1
Accidente o traumatismo grave	1) sí 2) no.	0 1
Intervención quirúrgica	1) sí 2) no.	0 1
Fiebres altas en el último año	1) hace menos de tres meses 2) hace más de tres meses 3) no.	-1 0 1
Frecuencia de consumo de alcohol	1) varias veces al día 2) todos los días 3) varias veces a la semana 4) una vez a la semana 5) casi nunca o nunca	0 1
Hábito de fumar	1) nunca 2) ocasionalmente 3) a diario	-1 0 1
Número de horas que pasa sentado al día	1-6	0 (sí) 1 (no)
Salida	Diagnóstico normal Alterado	N O

### 3.1. MATLAB

Se utilizó el software MATLAB como herramienta para los análisis de las muestras y la generación de la matriz de confusión, debido a su capacidad para procesar datos mediante una amplia variedad de herramientas y algoritmos.

En términos de análisis de datos, MATLAB ofrece una serie de herramientas para importar, procesar, visualizar y modelar datos. Por ejemplo, se pueden importar datos desde una variedad de formatos, como archivos de texto, hojas de cálculo, bases de datos, entre otros. Además, ofrece una amplia gama de herramientas para visualizar y explorar datos, como gráficos en 2D y 3D, histogramas, curvas de tendencia, entre otros.

También incluye una amplia gama de algoritmos de aprendizaje automático y modelado para el análisis de datos. Estos incluyen regresión lineal, regresión logística, árboles de decisión, redes neuronales y más. Estos algoritmos se integran con una variedad de herramientas para evaluar y validar los modelos, lo que permite a los usuarios evaluar la precisión y fiabilidad de los resultados [15].

### 3.2. Árboles de Decisión

Como técnica para el análisis de datos se escogieron los árboles de decisión, un método de aprendizaje supervisado que se utiliza para modelar y predecir valores numéricos o categóricos en un conjunto de datos. El objetivo es crear un modelo que pueda realizar predicciones sobre los datos de entrada, basándose en el historial de los datos.

La estructura básica de un árbol de decisión es un grafo que consiste en nodos y ramas. Cada nodo representa una decisión o una pregunta sobre los datos, mientras que cada rama representa una posible respuesta a la pregunta. La pregunta más relevante para predecir el valor deseado se encuentra en la raíz del árbol, mientras que las hojas contienen las predicciones finales [16].

#### 3.2.1. Algoritmo CART

Dentro de las características que nos ofrece el método de árboles de decisión ofrecido por MATLAB se encuentra el uso del Algoritmo *CART* una técnica de aprendizaje automático de árbol de decisión no paramétrico. Dependiendo de si la variable dependiente es numérica o categórica, produce árboles de clasificación o regresión respectivamente.

Los nodos en un árbol de decisión solo se pueden dividir en dos grupos, que se llama una división binaria, el algoritmo utiliza el índice de Gini para medir la impureza y seleccionar el atributo con la mayor reducción de impurezas para dividir los registros en un nodo y acepta valores de atributos faltantes [17]. Para calcular dicho índice se necesita la Fórmula (1):

$$gini = \sum_k p_k(1 - p_k) \quad (1)$$

que se traduce como la suma, para todas las clases, el producto de su proporción por 1 menos su proporción, donde un índice de Gini igual a cero significa que los datos en un nodo pertenecen a una sola categoría, lo que se conoce como un nodo puro, por otro lado, cuando es mayor que cero y puede llegar hasta uno, significa que los datos en el nodo provienen de más de una categoría, lo que se considera un nodo impuro.

En pocas palabras, esta característica de implementación se utilizó para medir la impureza de las características de la base de datos *Fertility* y así generar un árbol de decisión con únicamente los atributos más determinantes para la clasificación.

### 3.3. Evaluación Cruzada

Parte de este proyecto es realizar un test para el algoritmo de clasificación usado, siendo elegido el método de validación cruzada, una técnica de evaluación de modelos de aprendizaje automático utilizada para estimar la capacidad generalizada de un modelo.

En este método, se divide el conjunto de datos en subconjuntos, luego se entrena el modelo repetidamente usando diferentes subconjuntos como datos de entrenamiento y se evalúa el rendimiento del modelo en el subconjunto restante. Esto se hace de manera iterativa, hasta que todos los subconjuntos hayan sido utilizados como datos de prueba.

Finalmente, se promedian los resultados de todas las evaluaciones para obtener una estimación más precisa del rendimiento general del modelo. La validación cruzada se utiliza para evitar el sobreajuste y para mejorar la estimación de la capacidad de generalización de un modelo [18].

### 3.4. Matriz de Confusión

Después de realizar el entrenamiento y la validación se podrá generar una herramienta conocida como matriz de confusión utilizada para evaluar el rendimiento de un modelo de clasificación en el aprendizaje automático. Representa de manera gráfica la cantidad de veces que el modelo clasificó correctamente o incorrectamente a cada una de las categorías [19].

Se representa en forma de tabla cuadrada con filas y columnas que corresponden a las diferentes categorías o clases que se están clasificando y teniendo 4 posibles resultados a mostrar (Figura 1):

- Verdadero positivo o True Positive (TP): se refiere a una situación en la que el valor real es positivo y la prueba también lo predijo como positivo.
- Verdadero negativo o True Negative (TN): se refiere a una situación en la que el valor real es negativo y la prueba también lo predijo como negativo.
- Falso negativo o False Positive (FP): se refiere a una situación en la que el valor real es positivo, pero la prueba lo predijo como negativo. Esto se conoce como error tipo II.
- Falso positivo o False Negative (FN): se refiere a una situación en la que el valor real es negativo, pero la prueba lo predijo como positivo. Esto se conoce como error tipo I.

Valores Predicción	TP	FP	Clase 1
	FN	TN	Clase 2
	Valores Reales		

**Figura 1.** Representación Matriz de Confusión.

A partir de esta matriz se pueden calcular métricas que sirven de mucho para evaluar el modelo de aprendizaje automático utilizado.

#### 3.4.1. Precision (Precisión) y Accuracy (Exactitud)

La precisión es una medida utilizada para evaluar la calidad del modelo de aprendizaje automático en tareas de clasificación. La Fórmula (2) es utilizada para calcular la precisión:

$$precision = \frac{TP}{TP + FP} \quad (2)$$

La *accuracy* hace referencia a la consistencia de los valores obtenidos de una medida repetida de una cantidad. Una menor dispersión indica una mayor precisión. Su cálculo se hace mediante la Fórmula (3):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

#### 3.4.2. Recall (Exhaustividad) y Especificidad (Specificity)

Son dos medidas que nos muestran la habilidad del estimador para separar los casos positivos de los negativos. La sensibilidad es el porcentaje de resultados positivos reales, mientras que la especificidad es el porcentaje de resultados negativos reales. La primera, también conocida como Tasa de Verdaderos Positivos (*True Positive Rates* – TPR) proporciona información sobre la capacidad del modelo de aprendizaje automático para identificar datos. Su Fórmula se representa como (4):

$$recall = \frac{TP}{TP + FN} \quad (4)$$

Por otra parte, la segunda también conocida como la Tasa de Verdaderos Negativos (*True Negative Rate* – TNR). Se refiere a los casos negativos que el algoritmo ha clasificado correctamente, ver Fórmula (5).

$$specificity = \frac{TN}{TN + FP} \quad (5)$$

#### 3.4.4. Otras métricas de la matriz de confusión

Para la evaluación de algoritmos de aprendizaje automático también existen otras dos métricas que se conocen como Tasa de Falsos Positivos (*False Positive Rates* - FPR) y la Tasa de Falsos Negativos (*False Negative Rates* - FNR). La primera se refiere a la tasa de casos negativos que fueron equivocadamente clasificados como positivos por la Fórmula (6):

$$FPR = \frac{FP}{TN + FP} \quad (6)$$

Por otro lado, la segunda se refiere a la tasa de casos positivos que han sido clasificados de manera incorrecta según la Fórmula (7):

$$FPR = \frac{FN}{FN + TP} \quad (7)$$

#### 4. Resultados y Discusiones

En la primera etapa de la implementación se llevó a cabo la preparación de los datos, decidiendo que el 90% se usarían para el entrenamiento y la evaluación, mientras que el 10% restante se destinaría a las pruebas finales. Aplicando las herramientas disponibles en el ToolBox de MATLAB *Classification Learner*, se construyó un modelo representativo de un árbol de decisión (Figura 2). En dicho modelo, se utilizó el índice de *Gini* para reducir el número de atributos necesarios para clasificar la fertilidad, empleando los atributos contenidos en la base de datos. Debido a que otros métodos de clasificación requerían 9 atributos para clasificar la fertilidad, en este caso de uso se creó un árbol de decisión que solo utilizó 5 atributos. Se seleccionaron aquellos atributos con menor impureza, lo que sugiere que la clasificación de fertilidad puede lograrse utilizando solamente estos factores.

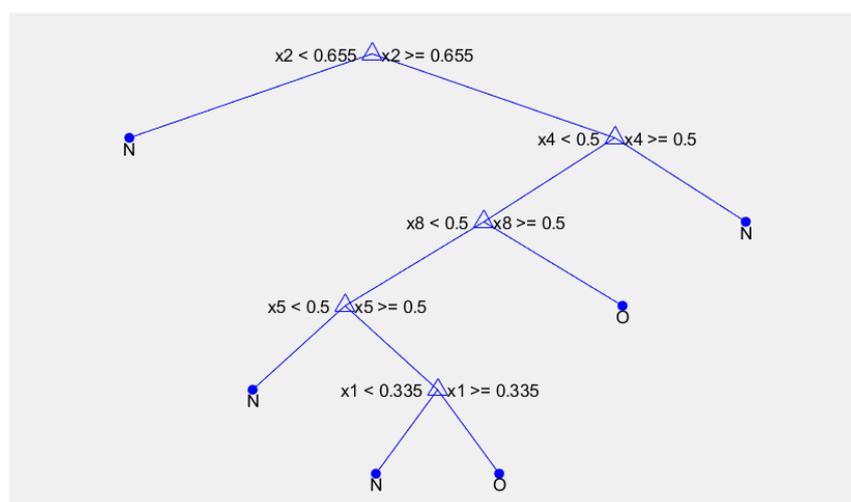
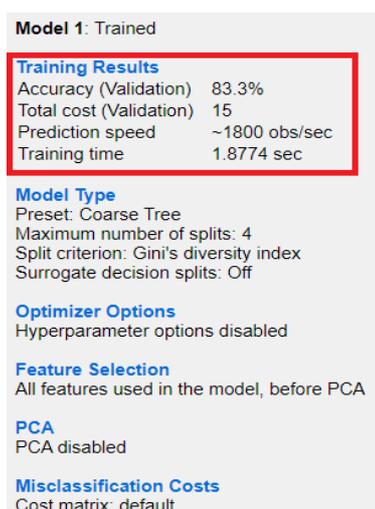


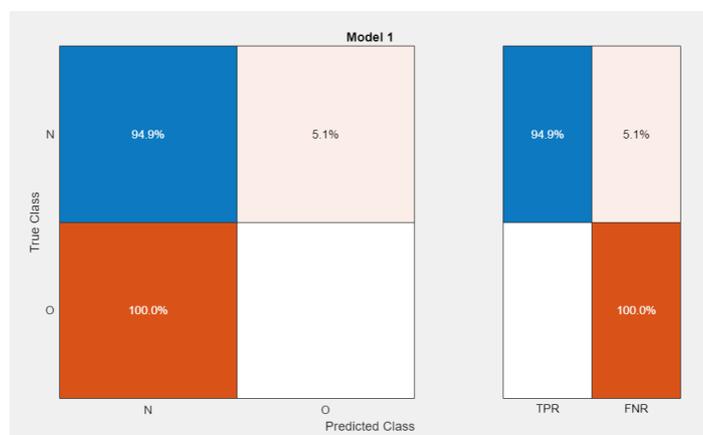
Figura 2. Árbol de decisión creado.

Para la segunda etapa correspondiente al entrenamiento y evaluación cruzada, se aplicó una división de los datos de entrenamiento en 5 subconjuntos, lo que permitió realizar esta etapa en 5 iteraciones. Durante el proceso, el Toolbox ofreció información interesante (Figura 3). El tiempo de entrenamiento fue de 1.8774 segundos, lo cual se considera muy rápido. Sin embargo, este tiempo se debió a la poca cantidad de muestras disponibles en la base de datos. Se recuerda que se ocupó el 90% de las 100 muestras totales para esta etapa. Esta cantidad de muestras afectó también al porcentaje de precisión en la validación, que fue de 83.3%. Aunque este porcentaje es considerado bueno, se tiene la certeza de que con un conjunto mayor de muestras se podría mejorar este valor hasta alcanzar una precisión mayor al 90%. En otros aspectos, se puede observar que la velocidad de predicción puede alcanzar las 1800 observaciones por segundo, lo cual es una velocidad aceptable cuando se requiera procesar un mayor número de muestras.



**Figura 3.** Información del entrenamiento.

Lo siguiente a analizar de esta etapa, es la matriz de confusión generada (Figura 4), la cual contiene la TPR y FNR para las dos clases que se manejan en este clasificador: Diagnóstico Normal (N) y Diagnóstico Alterado (O). En primer lugar, se observa que para la clase de O, la TPR no muestra ningún valor, lo que significa que nuestro árbol de decisión no cometió ningún error al clasificar las muestras utilizadas en el entrenamiento y la validación. Esto da una FNR del 100%, lo cual es justamente lo que se busca en un clasificador. Por el contrario, para la clase de N, el modelo arrojó una TPR del 94.9%, lo que indica que tuvo una buena evaluación de los datos y superó el umbral del 90% que es lo deseable. Sin embargo, no se logró evitar que un 5.1% de los datos se clasificara incorrectamente como falsos negativos. A pesar de ello, la matriz de confusión indica que el desempeño del clasificador fue bueno durante la etapa de entrenamiento y evaluación.



**Figura 4.** Matriz de confusión validación.

En la última etapa, aplicó un test utilizando el 10% de las muestras, como se mencionó anteriormente. En este caso, se obtuvo un porcentaje del 80% de aciertos en la clasificación y un costo de 2 (Figura 5). Aunque estos son números buenos, ciertamente se podrían mejorar, como se mencionó anteriormente, si se tuviera una base de datos con un mayor número de muestras a utilizar.

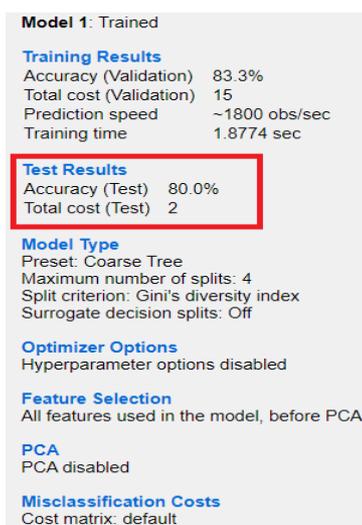


Figura 5. Resultados de la prueba.

También fue generada una matriz de confusión (Figura 6) en la cual se puede analizar que para la clase O se obtuvo una FNR del 100%, tal como se preveía con base en los resultados de la matriz de confusión en la etapa de entrenamiento y validación (Figura 4). En la clase N, se esperaba tener falsos negativos debido a los resultados previos. Es evidente que el 11.1% de las muestras fueron clasificadas de esa manera, lo que sugiere que esta clase requiere un mejor ajuste para reducir el porcentaje de FNR.

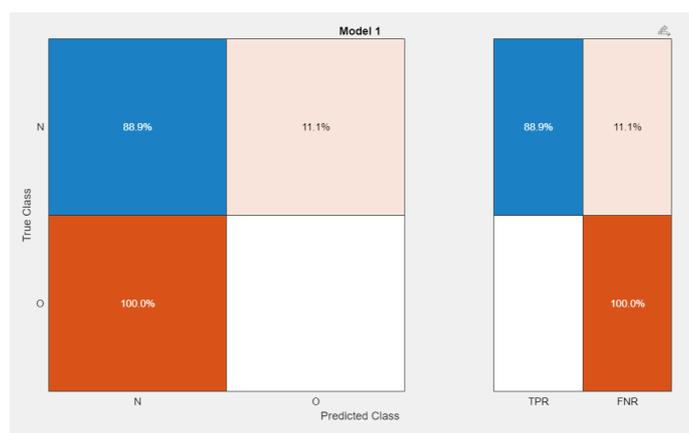


Figura 6. Matriz de confusión *test*.

Con estos resultados, se puede mencionar que este árbol de decisión creado con un menor número de atributos, gracias al índice de *Gini*, logra tener buen comportamiento al analizar muestras de fertilidad. Sin embargo, sería una gran idea recabar un mayor número de muestras para tener un entrenamiento que permita acercarse a un porcentaje de precisión igual o mayor al 90%.

## 5. Conclusiones

Se aplicó la técnica de árboles de decisión al 90% de las muestras de la base de datos *Fertility*. De esta manera, nos permitió identificar rápidamente las variables más relevantes que tienen impacto en la fertilidad de una persona. Como resultado, se obtuvieron cinco atributos principales: Edad, Accidente, Hábito de fumar, Intervención quirúrgica y Estación del año.

Gracias al motor de MATLAB, se pudieron obtener las diferentes gráficas mostradas en este trabajo. Estas describen y muestran de forma detallada los resultados obtenidos, proporcionando una visión general de cómo se clasifican los datos y cómo el sistema los utiliza para interpretar los resultados.

Como se puede apreciar en este trabajo, se obtuvo un nivel de precisión del 83.3%, el cual resulta aceptable para la tarea de clasificación. Sin embargo, es importante destacar que este valor puede variar significativamente en función de los métodos de validación utilizados, el porcentaje de datos de pruebas, el porcentaje de validación y el número de atributos empleados.

El número de datos en la base de datos utilizada es relativamente pequeño, por lo que los resultados del entrenamiento se vieron afectados por la falta de información para clasificar correctamente los diferentes tipos de muestras. Se espera que, al contar con un mayor número de registros de muestras de esperma, se logre aumentar la precisión de las clasificaciones a medida que aumente el grado de concentración de los datos recopilados.

Todos los métodos de clasificación son susceptibles de mejora. La clave es ajustar los parámetros y los valores de los porcentajes de datos de entrenamiento, validación y pruebas y observar cuáles valores se aplican mejor al problema existente para incrementar el porcentaje de éxito. Es importante recordar que no existe un valor específico para clasificar los datos, pero es necesario tener cuidado con el sobreajuste (*overfitting*), ya que podría sobreentrenar la red y los valores obtenidos podrían ser mal clasificados.

Si se requieren datos más precisos, es necesario realizar pruebas con diferentes técnicas, analizar y comparar qué modelo de clasificación se adapta mejor al problema existente, dependiendo de los datos con los que se cuenten y los métodos aplicados. Es importante buscar aquel con el mayor porcentaje de éxito.

Algunos trabajos similares han utilizado el software WEKA para comparar varios métodos de clasificación. Sin embargo, no indican el porcentaje de datos utilizados para el entrenamiento, la validación y pruebas, por lo que deja al lector con falta de información para comprender mejor lo que se ha realizado. Por lo tanto, sería recomendable implementar otros métodos de clasificación utilizando el software de MATLAB, que permite un análisis más detallado de los resultados obtenidos.

## 7. Referencias

- [1] Mundial BIRF AIF. (2020). *Tasa de fertilidad, total (nacimientos por cada mujer)*. DataBank. <https://datos.bancomundial.org/indicador/SP.DYN.TFRT.IN>
- [2] World Health Organization. (2020). *Infertility*. <https://www.who.int/news-room/fact-sheets/detail/infertility>
- [3] Criado, M. A. (2022). *La calidad del esperma de los humanos ha bajado a la mitad en el último medio siglo*. El País. <https://elpais.com/ciencia/2022-11-15/la-calidad-del-esperma-de-los-humanos-ha-bajado-a-la-mitad-en-el-ultimo-medio-siglo.html>
- [4] Castrillón, O. D., Arango, J. A., Castillo, L. F. (2022). Análisis de la fertilidad por medio de técnicas de minería de datos. *Información tecnológica*, 33 (3), 203-212. <https://dx.doi.org/10.4067/S0718-07642022000300203>
- [5] Velilla-Hernández, E., Velilla-Hernández, P. A., Cardona-Maya, W. (2013). Aproximación al desempeño operativo de un modelo de redes neuronales en el diagnóstico de la infertilidad masculina. *Revista Colombiana de Obstetricia y Ginecología*, 64 (3), 222-228. <https://doi.org/10.18597/rcog.103>
- [6] Causa-Andrieu, P., Couture, J. M., Benítez-Mendes, A. C., Napoli, N., Chacón, C. R. B., Ulla, M. (2017). Infertilidad en la mujer: importancia del análisis combinado de los métodos por imágenes. *Reporte de casos. Ginecología y obstetricia de México*, 85 (12), 846-852. <https://doi.org/10.24245/gom.v85i12.1442>
- [7] Mendoza, F., De la Hoz Manotas, A., Ariza, P., Sepulveda Ojeda, J., Melo, M. (2016). Fertility Analysis Method Based on Supervised and Unsupervised Data Mining *Techniques. International Journal of Applied Engineering*, 11 (21), 10374-10379.
- [8] Milewska, A., Jankowska, D., Cwalina, U., Citko, D., Więsak, T., Acacio, B., Milewski, R. (2017). Prediction of Infertility Treatment Outcomes Using Classification Trees. *Studies in Logic, Grammar and Rhetoric*, 47 (1), 7-19. <https://doi.org/10.1515/slgr-2016-0043>
- [9] Solarte Martínez, G. R., Soto Mejía, J. A. (2011). Árboles de decisiones en el diagnóstico de enfermedades cardiovasculares. *Scientia Et Technica*, 16 (49), 104-109.
- [10] Pérez López, C., Santin González, D. (2007). *Minería de datos. Técnicas y herramientas: técnicas y herramientas*. Ediciones Paraninfo, S. A.
- [11] Gil, D., Girela, J. L., De Juan, J., Gomez-Torres, J. J., Johnsson, M. (2012). Predicting seminal quality with artificial intelligence methods. *Expert Systems with Applications*, 39 (16), 12564-12573. <https://doi.org/10.1016/j.eswa.2012.05.028>

- [12]del Castillo Collazo, N. (2020). Incidencia en el pronóstico al aplicar reducción de variables. Un ejemplo práctico. *Revista de Investigación en Tecnologías de la Información (RITI)*, 8 (15), 50–69.  
<https://doi.org/10.36825/RITI.08.15.006>
- [13]Corrales Gasca, P. J., Razo Granados, J. A., Violante Hernández, M. Á., Rodríguez Hernández, I. J., Ramírez Soto, R. (2015). Métodos de clasificación: Análisis de fertilidad. *Pistas Educativas*, 35 (111), 43-57.
- [14]Hernández Figueroa, L. J., Serna Manríquez, A., Gómez Melesio, J. A., Hurtado Mendoza, J. G., Ramírez Pérez, N. V. (2015). Análisis de métodos de clasificación para el diagnóstico de fertilidad. *Pistas Educativas*, 36 (114), 243-259. <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/301>
- [15]MathWorks. (2021). *MATLAB*. <https://www.mathworks.com/products/matlab.html>
- [16]Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1 (1), 81-106.  
<https://doi.org/10.1007/BF00116251>
- [17]Lewis, R. J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*. Annual Meeting of the Society for Academic Emergency Medicine. San Francisco, California.
- [18]Chollet, F. (2018). *Deep learning with Python*. Manning Publications.
- [19]Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2 (1), 37-63.  
<https://doi.org/10.48550/arXiv.2010.16061>