

Análisis de clúster para la clasificación de pacientes atendidos en consulta de infertilidad

Cluster analysis for the classification of patients attended at the infertility consultation

Mabel A Rodríguez-Monteagudo¹

0000-0001-8681-6083

Raidell Avello-Martínez²

0000-0001-7200-632X

Amanda Avello-Rodríguez¹

0000-0002-3336-6450

¹ Universidad de Ciencias Médicas de Cienfuegos. Cuba.

² Universidad Bolivariana del Ecuador. Ecuador.¹¹

Autor para la correspondencia: ravellom@ube.edu.ec

RESUMEN

Introducción: el espermograma es el examen clínico que brinda la visión más amplia de la capacidad reproductiva del hombre e incluye un gran número de parámetros a interpretar.

Objetivo: caracterizar a los pacientes masculinos de la Consulta de Atención a la Pareja Infértil de Cienfuegos mediante el análisis de clúster de las variables del espermograma.

Métodos: se llevó a cabo un estudio transversal de los espermogramas realizados a pacientes de la Consulta de Atención a la Pareja Infértil del municipio de Cienfuegos, en el período 2019 – 2020. La muestra estuvo compuesta por 128 pacientes. Se crearon 2 clústeres utilizando el algoritmo k-medias jerárquico, ejecutado en el lenguaje estadístico R.

Resultados: los valores del espermograma arrojaron rangos muy amplios en todas las variables. El primer grupo, con valores más favorables en los parámetros del espermograma, incluyó a un 82,8 % de los pacientes. Existieron diferencias significativas entre los dos clústeres para todas las variables, excepto leucocitos, y las que resultaron más definitorias para la formación de los mismos fueron la vitalidad, movilidad rápida, movilidad total y concentración.

Conclusiones: se consiguió agrupar a los pacientes de la consulta de infertilidad de Cienfuegos en dos clústeres con diferencias significativas en la mayoría de los parámetros del espermograma, lo que permitió caracterizar globalmente a esta población. Estos clústeres pudieran servir posteriormente para crear una herramienta complementaria para el diagnóstico individual de estos pacientes.

Palabras clave: análisis de clúster; infertilidad; espermograma; k-medias jerárquico; aprendizaje no supervisado.



ABSTRACT

Introduction: Spermogram is the clinical examination that provides the broadest view of the male reproductive capacity and includes a large number of parameters to be interpreted.

Objective: To characterize the male patients attended at the Consultation for the Care of the Infertile Couple in Cienfuegos through a cluster analysis of spermogram variables.

Methods: A cross-sectional study of the spermograms performed on patients seen at the Consultation for the Care of the Infertile Couple in Cienfuegos Municipality was carried out in the period 2019 - 2020. The sample consisted of 128 patients. Two clusters were created using the hierarchical k-means algorithm, executed by the R statistical language.

Results: The values of the spermogram showed very wide ranges in all the variables. The first group, with more favorable values in the spermogram parameters, included 82.8% of the patients. There were significant differences between the two clusters for all the variables, except leukocytes; the ones that were most defining for their formation were: vitality, rapid mobility, total mobility, and concentration.

Conclusions: It was possible to group the patients from the infertility consultation in Cienfuegos into two clusters with significant differences in most of the spermogram parameters, which made it possible to characterize this population globally. These clusters could later be used to create a complementary tool for the individual diagnosis of these patients.

Keywords: cluster analysis; infertility; spermogram; hierarchical k-means; unsupervised learning.

Recibido: 24/09/2023

Aprobado: 20/02/2024

Introducción

Desde la antigüedad, la incapacidad para generar descendencia ha constituido un problema con consecuencias negativas, sobre todo en las esferas psíquica y social. Las estadísticas referentes a la prevalencia de la imposibilidad de concebir no son exactas: se estima por la Organización Mundial de la Salud (OMS) que aproximadamente entre el 10 y el 20 % de las parejas experimentan alguna imposibilidad de concebir durante su vida reproductiva. ⁽¹⁾ "Esto supone que aproximadamente una de cada siete parejas en edad reproductiva presentará problemas para tener descendencia." ⁽²⁾

La infertilidad es conceptualizada como la incapacidad de concebir cuando se ha cumplido 1 año de relaciones sexuales heterólogas sin protección. Muchas veces se emplean los términos esterilidad e infertilidad indistintamente, mientras que en otras ocasiones se asocia la infertilidad a la pareja que no consigue la gestación, y esterilidad a la imposibilidad absoluta de concebir. ^(1, 3)



Se estima que más de 70 millones de parejas presentan problemas de infertilidad a escala global y que alrededor de un millón requieren técnicas de reproducción asistida. En Cuba, en torno a las 300 000 parejas se consideran susceptibles de atención en consulta de infertilidad, con una incidencia entre 12 y 15 %.^(4, 5)

“Recientes estudios señalan que el aporte masculino y femenino son equivalentemente importantes en la contribución a un embrión totipotente, que pueda diferenciarse con éxito en un individuo sano.”⁽⁶⁾ “Se plantea a su vez que el factor masculino está afectado en alrededor del 50 % de las parejas infértiles.”⁽⁴⁾

La fertilidad masculina es un área de la investigación médica que está ganando cada vez más atención debido a varios estudios que indican una disminución global en la calidad del semen durante las últimas décadas, atribuida muchas veces a diferencias geográficas.^(7, 8) Por tanto, continúa siendo necesario actualizar y perfeccionar constantemente los métodos empleados en el diagnóstico y tratamiento de dicho trastorno.

El espermograma continúa siendo la prueba de elección para evaluar la fertilidad masculina, por ser la que brinda la visión más amplia de la capacidad reproductiva del hombre.^(9, 10) Es un examen sencillo, de bajo costo, que permite enfocar el tratamiento de la infertilidad hacia determinadas aristas, así como evaluar progresivamente los resultados de dicho tratamiento, ya sea clínico o quirúrgico.⁽¹¹⁻¹³⁾

Los estudios que relacionan la calidad de la eyaculación y la fertilidad han existido desde la década de 1930.⁽¹⁴⁾ En la década de los 80s, un grupo de científicos del área de la medicina, seleccionados por la OMS, confeccionaron un manual donde se describían los valores normales del espermograma y se establecían pautas para el análisis del semen en el laboratorio. Desde entonces, han actualizado el manual cuatro veces, publicándose la última versión en 2010.⁽¹⁵⁾

El análisis seminal incluye tanto la evaluación macroscópica donde se valoran pH, volumen, licuefacción y viscosidad, como la observación microscópica donde movilidad, viabilidad, concentración y morfología son valorados.⁽¹¹⁻¹³⁾ “Las características del semen varían según la edad, estado nutricional, patologías asociadas, exposición a diversos contaminantes del medio ambiente, temperatura corporal y ambiental, lugar de residencia, dieta y consumo de alcohol y tabaco.”⁽¹¹⁾

La interpretación del espermograma es compleja debido al gran número de parámetros que analiza y la subjetividad que conlleva su integración y relación con las características clínicas del paciente. Por tal motivo puede resultar de utilidad el uso de herramientas digitales de aprendizaje computacional, como la generación de clústeres, para caracterizar conjuntos de pacientes estudiados por infertilidad o para complementar su diagnóstico individual, y facilitar la toma de decisiones sobre políticas de salud, en el primer caso, o sobre tratamientos a indicar, en el segundo.



El análisis de clúster consiste en aglomerar los datos en grupos, donde cada grupo agrega datos con características similares dentro de sí, y a su vez diferentes entre ellos. (16-19) Estos clústeres pueden generarse de forma automática (aprendizaje no supervisado) según lo sugiera el análisis de los datos, o pueden disponerse inicialmente por el programador (aprendizaje supervisado).^(20, 21) La herramienta, luego de ser entrenada, es capaz de asignar los nuevos pacientes automáticamente a uno de los clústeres, e incluso volver a procesar los datos y reorganizar los grupos, lo que podría devenir un eficiente sistema de apoyo al médico para el análisis global del espermograma y del paciente en general. Además, al caracterizar grupos de pacientes, los clústeres pueden sugerir cuáles variables resultan más importantes en la diferenciación de los pacientes, y de esta manera, ampliar la comprensión sobre la situación de salud de los mismos.

El objetivo de este trabajo fue caracterizar a los pacientes masculinos de la Consulta de Atención a la Pareja Infértil de Cienfuegos mediante el análisis de clúster de las variables del espermograma.

Métodos

Se llevó a cabo un estudio transversal de los espermogramas realizados a pacientes de la Consulta de Atención a la Pareja Infértil del municipio Cienfuegos en el período 2019 – 2020. La edad de los participantes estuvo entre 22 y 58 años. Se excluyeron del estudio 12 pacientes cuyos registros presentaban datos incompletos. Por último, la muestra estuvo compuesta por 128 pacientes. El análisis espermático se realizó en los laboratorios de los Policlínicos de las Área de Salud II y VII del municipio Cienfuegos y del Centro de Reproducción Asistida Regional.

Para el análisis de los datos se emplearon las variables volumen, concentración, movilidad rápida, movilidad total (movilidad rápida + movilidad lenta), morfología, vitalidad y leucocitos, de las cuales inicialmente se obtuvieron la media, mediana, desviación estándar, y límites inferior y superior.

Todo el procesamiento de los datos fue realizado con paquetes del lenguaje de programación estadístico R.⁽²²⁾ Se empleó el análisis de clúster, mediante aprendizaje computacional no supervisado. No era de interés el establecimiento previo de los grupos pues limitaba las posibilidades de análisis de la presente muestra.

Antes de proceder con el análisis de clúster se estandarizaron las variables, o sea, se llevaron a media 0 - desviación estándar 1. Además, se realizó un análisis gráfico de aglomeraciones para el set de datos, donde se apreciaron algunos núcleos de datos agrupados, lo cual sugirió que podía aplicarse un análisis de clúster. Adicionalmente, se calculó el estadístico Hopkins,⁽²³⁾ resultando 0.260, un valor medio (mientras más cercano a 0 es mayor la posibilidad de encontrar agrupamientos y mientras más cercano a 0.5, menor).

Se realizó una reducción de dimensiones usando un Análisis de Componentes Principales y se realizó un gráfico de dispersión entre las dimensiones 1 y 2. La dispersión sugirió la posibilidad



de crear al menos dos clústeres. También se realizó un análisis para encontrar el número óptimo de clústeres con el paquete *factoextra*, el cual sugirió que el número óptimo podía ser 2 o 4 clústeres. Además, los resultados también confirmaron que la mayoría de los algoritmos (k-medias, pam, jerárquico, fuzzy k-medias, entre otros) sugerían 2 o 4 clústeres.

Por razones prácticas, se consideró más conveniente utilizar 2 clústeres, de manera que fuera posible una delimitación, en el contexto médico, entre pacientes con mayores y menores posibilidades de lograr un embarazo. Además, el uso de 4 clústeres podía resultar engorroso y carecía de argumentos que justificaran su necesidad.

Los resultados de la función *cIValid()* sugirieron, además, que los algoritmos óptimos en este caso eran k-medias y árbol jerárquico. Sin embargo, los estudios más actuales proponen el uso del algoritmo k-medias jerárquico, que es una fusión de k-medias con un preprocesamiento jerárquico de los datos para perfeccionar el posicionamiento inicial de los centroides, que es la principal debilidad del k-medias.⁽²⁴⁻²⁷⁾ A su vez, el uso del k-medias jerárquico no entra en contradicción con los resultados de la función *cIValid()* puesto que toma los elementos más apropiados de los dos algoritmos propuestos. Para aplicar el k-medias jerárquico, se utilizó el paquete *clustertend* y para la visualización de los clústeres, el paquete *factoextra*.

Además, se recalcularon los clústeres utilizando los algoritmos k-medias y árbol jerárquico por separado con el objetivo de validar los resultados. El índice Rand ajustado (adjusted Rand index, ARI) resultó 0.86 para el k-medias y 0.79 para el árbol jerárquico aproximadamente, ambos cercanos a 1, por lo que hay alta correspondencia entre los resultados de cada uno de estos algoritmos y el k-medias jerárquico. Para confirmar los resultados, se aplicó la prueba t de student de diferencia de medias en todas las variables.

Por otro lado, se calcularon los indicadores MDA (Mean Decrease Accuracy, Disminución media de la precisión) y MDG (Mean Decrease Gini, Disminución media de Gini), aplicando el algoritmo de aprendizaje automático supervisado *RandomForest*, que miden la importancia de cada variable en la creación de los clústeres.⁽²⁸⁾

Los principios éticos fueron respetados en todo momento de la investigación. Para el análisis y publicación de los datos se omitieron los nombres y dirección particular de los pacientes, con el objetivo de respetar su privacidad. Además, se obtuvo el consentimiento informado de todos los pacientes.

Resultados

En el análisis de los valores de los parámetros del espermograma obtenidos de la muestra utilizada (ver tabla 1) se observan rangos muy amplios, extendiéndose desde 0 % hasta más de un 80 % en las variables movilidad rápida, movilidad total, morfología y vitalidad, y desde 0 hasta 350 millones de espermatozoides por ml de esperma en cuanto a la concentración. Las medias de los valores de todos los parámetros se encuentran por encima de los límites inferiores propuestos por la OMS para el espermograma. Además, los límites en todas las variables son superados por más de un 50 % de los pacientes. Los parámetros más afectados



fueron la movilidad rápida (35,2 %) y la movilidad total (14,8 %), seguidas de la concentración (11,7 %). La media de los leucocitos encontrados en las muestras fue de 1 millón por ml.

Tabla 1- Medidas de tendencia central y dispersión de los resultados por parámetros del espermograma.

Medidas	Volumen (ml)	Concentración (millones por ml)	Movilidad Rápida (%)	Movilidad Total (%)	Morfología (%)	Vitalidad (%)	Leucocitos (millones por ml)
Media	4,24	89,1	37,4	57,7	46,3	71,8	1,5
Mediana	4,2	87,5	36,5	62,5	40,0	78,5	1,0
DE	1,31	48,6	20,2	1,9	21,6	24,8	0,1
Mínimo	1	0	0	0	0		0
Máximo	10	350	85	95	92		3
Valores OMS (2010)	≥ 1,5	≥ 39	≥ 32	≥ 40	≥ 4	≥ 58	-
% de cumplimiento	98,4	88,3	64,8	85,2	96,1	98	-

La figura 1 muestra la asignación de los individuos a los grupos y la disposición de los centroides. Para K = 2 grupos, el agrupamiento automático con k-medias jerárquico dividió la muestra en 106 sujetos (82,8 %) en el clúster 1 y 22 sujetos (17,2 %) en el clúster 2.

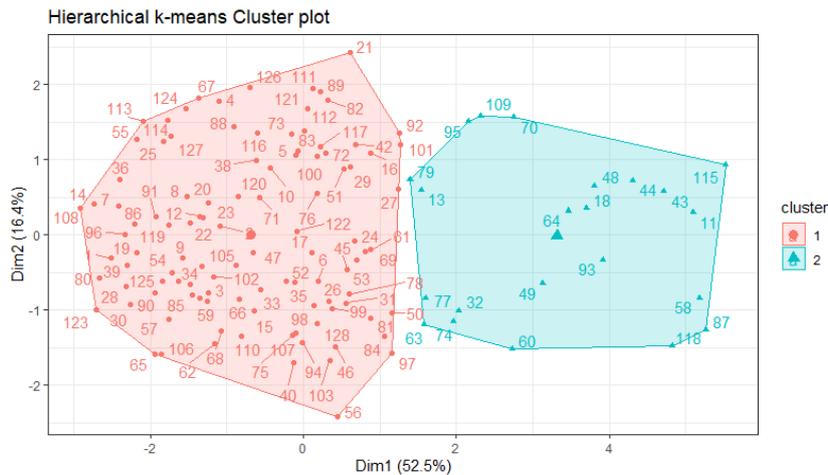


Fig. 1- Clasificación de los individuos en los clústeres 1 (polígono de color rojo) y 2 (polígono de color azul) mediante k-medias jerárquico.

En la tabla 2, se muestran las medias y desviaciones estándar de las variables en cada clúster. Realizando una comparación entre ambos, encontramos que el primero se caracteriza por valores más altos, o sea, más favorables hacia la fertilidad en todas las variables, excepto en



leucocitos, que tiene una dirección inversa, y presenta una media inferior para este clúster (1,43 en relación a 1,82 millones por ml).

Si se toman como punto de comparación los límites propuestos por la OMS ⁽¹⁵⁾ para los parámetros del espermograma, se obtiene que la media del volumen, la concentración, la movilidad rápida, la movilidad total, la morfología y la vitalidad estuvieron por encima de dichos umbrales en el caso del clúster 1, mientras que la variable leucocitos no es referida por dicha organización. Si se analiza el clúster 2 en este sentido, se encuentra que las variables concentración (30,41 millones por ml), movilidad rápida (5,73 %), movilidad total (19,14 %) y vitalidad (31,41 %) estuvieron por debajo de los límites de referencia de la OMS.

Además, los resultados que se observan en la tabla 2 evidencian que existen diferencias significativas entre las medias de ambos clústeres en todas las variables, siempre con p-valores < 0.001, menos en el caso del volumen, que de todas maneras fue bajo (0.006). Solo la variable leucocitos arrojó diferencias no significativas de la media ($p = 0.137$).

Tabla 2- Prueba t de Student entre los clústeres 1 y 2, por variables estudiadas.

	Clúster 1	Clúster 2	t	gl	p valor	Diferencia de medias
	M (SD)	M (SD)				
Volumen	4,38 (1,26)	3,54 (1,36)	2.80	2,80	0,006	0,838
Concentración	101,24 (42,60)	30,41 (29,10)	7.43	7,43	< ,001	70,827
Movilidad Rápida	43,97 (15,19)	5,73 (6,45)	11.57	11,57	< ,001	38,244
Movilidad Total	65,68 (12,68)	19,14 (14,95)	15.19	15,19	< ,001	46,543
Morfología	50,85 (19,43)	24,18 (18,16)	5.92	5,92	< ,001	26,667
Vitalidad	80,12 (14,34)	31,41 (25,43)	12.44	12,44	< ,001	48,714
Leucocitos	1,43 (1,06)	1,82 (1,26)	-1.50	-1,50	0,137	-0,384

Como se observa en la misma tabla, las variables vitalidad, movilidad rápida (MovRap), movilidad total (MovTotal) y concentración (Concentr) fueron las de mayores diferencias. Esto fue comprobado gráficamente utilizando diagramas de cajas y gráficos de líneas, con los valores estandarizados (Ver figura 2, a y b). Según los indicadores MDA y MDG, estas mismas 4 variables resultaron las más definitorias para la creación de los clústeres. (Ver figura 3)



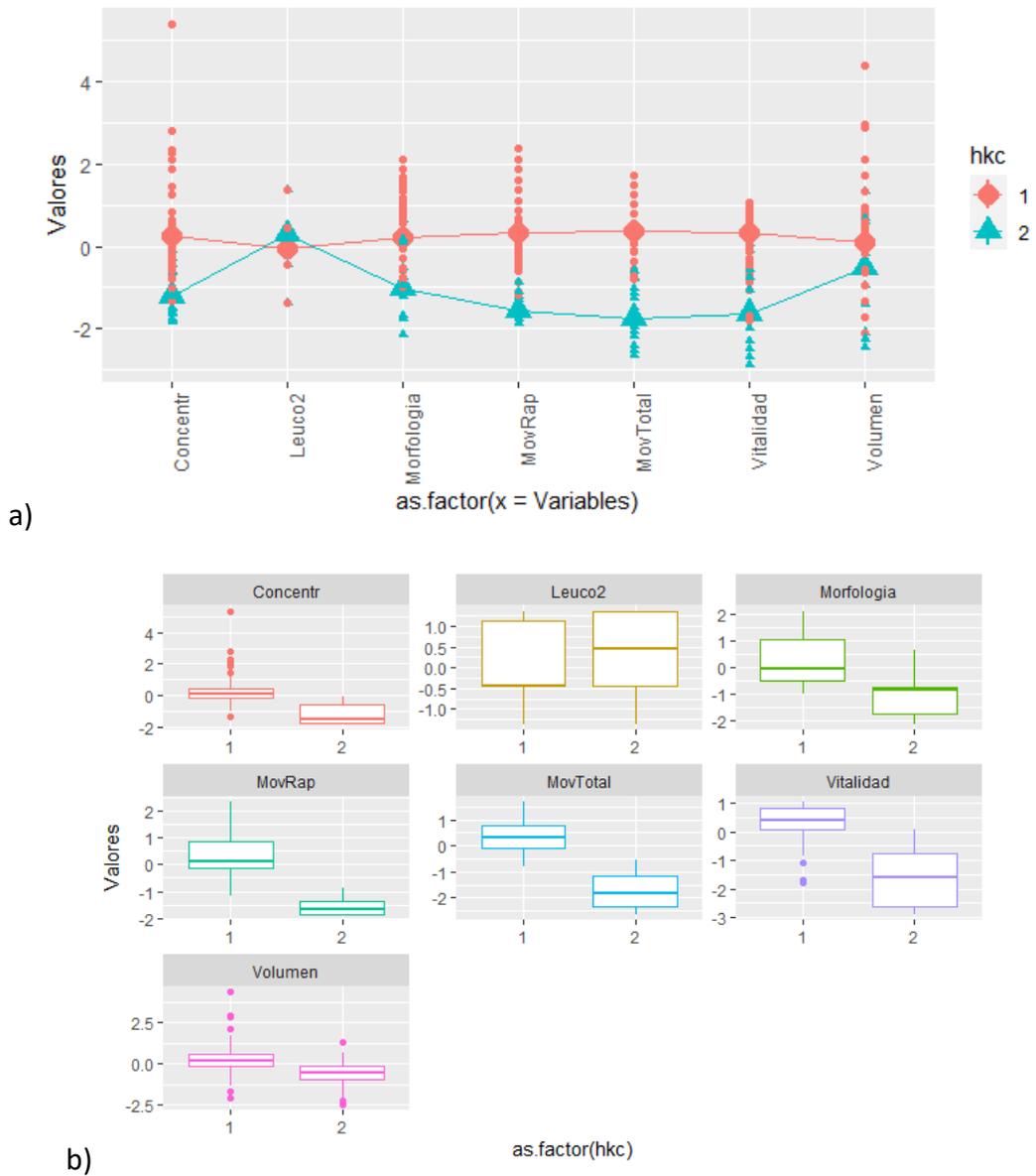


Fig. 2- a) gráfico de líneas y b) diagrama de cajas de las variables por cada grupo, con valores estandarizados.



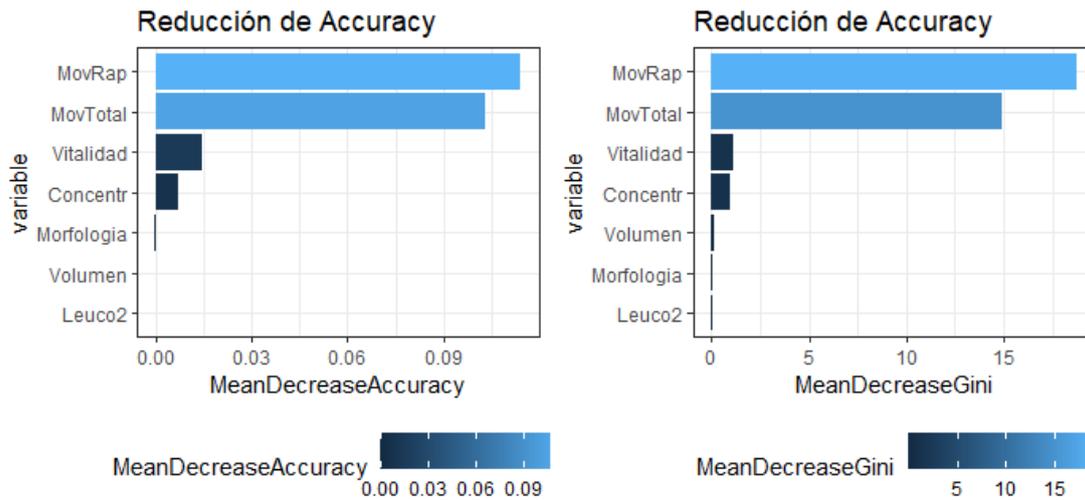


Fig. 3- Indicadores MDA y MDG de importancia de variables según algoritmo Random Forest.

Discusión

En la presente investigación se caracterizó una muestra de 128 pacientes masculinos atendidos en consulta de infertilidad. Se encontró una gran variabilidad en los resultados del espermograma entre estos y los parámetros más afectados fueron la movilidad rápida, la movilidad total y la concentración, aunque en general más de la mitad de los pacientes se encontraban dentro de la normalidad en todas las variables.

El aporte principal de esta investigación lo supone el empleo de algoritmos de aprendizaje automático, no supervisado, fundamentalmente el k-medias jerárquico, para el análisis de clúster de los 7 parámetros del espermograma, con lo que se logró clasificar a los pacientes atendidos en la consulta de infertilidad en dos clústeres o grupos con diferencias significativas entre ellos en los valores de casi todas las variables. Con respecto a este enfoque no se ha encontrado literatura precedente.

El primer clúster obtenido agrupa a más de un 80 % de los pacientes. Los mismos parecen tener mayores posibilidades de concebir pues presentan valores más elevados en todas las variables,^{(7), (8), (13), (16)} y la mayoría de estas superan los límites propuestos por la OMS.⁽¹⁵⁾ En el caso de la variable leucocitos, la dirección es inversa, presentándose en este clúster valores más bajos, que son más favorables. En el segundo clúster, con una quinta parte de los pacientes, se observa lo contrario, valores bajos en casi todos los parámetros, altos en los leucocitos y que estuvieron por debajo de lo referido por la OMS en 4 variables. Por tanto, presentan una tendencia mayor hacia la infertilidad.



Mediante estos análisis puede interpretarse que la mayor parte de los pacientes masculinos presenta una situación favorable respecto a la fertilidad. Por tanto, es posible que el componente femenino tenga mayor influencia en la afectación de la fertilidad entre las parejas atendidas en la consulta, y se deba hacer énfasis en el estudio de dichas féminas. En relación a esto, el estudio de Rojas et al. ⁽²⁹⁾ en la propia provincia de Cienfuegos devela un elevado porcentaje de contagios (42 %) por *Chlamydia trachomatis* entre las pacientes sometidas a técnicas de reproducción asistida en el Hospital Gustavo Aldereguía Lima, lo que sustenta la interpretación anterior.

Otro aporte muy interesante fue determinar, también mediante el uso de algoritmos de inteligencia artificial, ⁽³⁰⁾ qué parámetros del espermograma tuvieron la mayor influencia en la diferenciación de ambos grupos, o lo que es lo mismo, las variables que caracterizan a estos clústeres, las cuales fueron la vitalidad, la movilidad rápida, la movilidad total y la concentración. A su vez, dichas variables coinciden con las que han sido valoradas por la OMS como las más predisponentes a la infertilidad. ⁽³¹⁾ Puede interpretarse, por tanto, que estas condiciones deben ser atendidas con más urgencia en la consulta de infertilidad, para acercar las características de los pacientes del grupo más desfavorecido a las del otro.

En este sentido, un estudio de 82 pacientes masculinos de la Consulta de Infertilidad de las Tunas encontró que la alteración más común fue la astenozoospermia (92,7 %) seguida de la teratozoospermia (42,6 %) y la oligozoospermia (23,2 %). ⁽³²⁾ Estos porcentajes de afectación son mucho mayores que los encontrados en la presente muestra, pero algunos parámetros determinantes coinciden. Dichos autores comentan que la disminución de la motilidad espermática es una de las principales causas de infertilidad masculina y que puede tener su origen en anomalías en la morfología flagelar.

En contraste, en una investigación realizada en Marruecos con 482 pacientes atendidos por infertilidad en la pareja o individualmente por varicocele o ectopia testicular, se encontró que 53,1 % presentaron alteraciones del espermograma, siendo más frecuentemente afectada la vitalidad (36,9 %), seguida de la concentración espermática (29,7 %) y la morfología (29,3 %). Las variaciones en este caso se pueden justificar por la diferencia geográfica. ⁽³³⁾

La concentración de leucocitos no fue muy determinante para la diferenciación de los clústeres por el algoritmo. Aun así, es de resaltar la importancia del análisis de este factor en consulta, sobre todo porque las demás variables pueden estar afectadas en presencia de leucospermia, como dan a conocer varios autores. ⁽³⁴⁾ Además, los pacientes con infecciones requieren un tratamiento diferenciado y pueden presentar numerosas complicaciones. Se requiere, de cualquier forma, un análisis posterior de mayor profundidad en relación a este parámetro.

Por otro lado, algunos estudios ^{(10), (13)} han mostrado que el conteo total de espermatozoides móviles (volumen x concentración x movilidad) ha sido el factor más predictivo en la determinación de la fertilidad en contraste con el volumen, la concentración y la movilidad individualmente. Ello fundamenta en parte la importancia de la visión integradora de estas variables.



Será necesario un estudio posterior para determinar si estos clústeres pueden ser usados para mejorar el entendimiento de la condición de las parejas atendidas en consulta de infertilidad y el pronóstico de éxito en su tratamiento individual. A pesar de que el espermograma es la mejor herramienta con que se cuenta para valorar la fertilidad del hombre, no son la calidad y cantidad del semen los únicos factores predisponentes de la misma, si tenemos en cuenta que otros factores biopsicosociales son determinantes fundamentales. ⁽¹⁶⁾

El objetivo final de un algoritmo de este tipo sería la elaboración de un instrumento fácilmente manejable para el profesional de la salud que, mediante la introducción de los datos de un nuevo paciente, lo clasifique de forma automática a la vez que se retroalimenta el sistema. Se brindaría así, al médico de asistencia de la consulta de infertilidad, una herramienta complementaria útil a la hora de determinar la dirección del diagnóstico y tratamiento de la pareja.

Sin embargo, varias limitaciones se presentan en este análisis. La primera está dada porque la muestra la constituyeron únicamente pacientes de la consulta de infertilidad, mientras que una muestra representativa de la población en general sería más conveniente a la hora de crear una herramienta complementaria o diagnóstica global de este tipo. La segunda limitación para ello, es que no se calculó la asociación estadística de la alteración de los parámetros del espermograma con la edad, patologías asociadas y resultado alcanzado por los pacientes durante y después de la terapéutica, lo que sería fundamental para el acierto del algoritmo y tal vez una distinción más clara entre “fértiles” e “infértiles”.

En tercer lugar hay que tener en cuenta que las indicaciones a seguir por los pacientes en los días previos a la recogida de muestra, como la abstinencia sexual, el no consumo de bebidas alcohólicas y de tabaco, no uso de antibióticos, lavado de genitales y la micción pre-eyaculación, no siempre seguidas, pueden influir en la veracidad del resultado.⁽³⁾ Asimismo, el análisis manual del semen requiere personal de laboratorio capacitado, e incluso cuando se realiza de acuerdo con las directrices de la OMS, puede ser propenso a una gran variabilidad entre los diferentes laboratorios, por lo que los resultados presentados no deben extrapolarse a poblaciones no locales.

Conclusiones

Mediante el análisis de clúster de los parámetros del espermograma se agrupó a los pacientes de la consulta de infertilidad de Cienfuegos en dos grupos con diferencias significativas entre las medias en casi todas las variables, excepto leucocitos. El primer clúster incluyó a la mayor parte de los pacientes, que presentaron una situación favorable para la fertilidad; mientras que el segundo agrupa a una minoría con mayor alteración en los resultados del espermograma. Además, se logró determinar que las variables vitalidad, movilidad rápida, movilidad total y concentración revistieron la mayor importancia en la caracterización de los grupos y deben valorarse con más énfasis en consulta.



Referencias

1. Rigol Ricardo O, Santiesteban Alba SR. Obstetricia y Ginecología. 4 ed [Internet]. La Habana: Editorial Ciencias Médicas; 2023 [Citado 15/08/2023]. Disponible en: <http://www.bvscuba.sld.cu/libro/obstetricia-y-ginecologia-cuarta-edicion/>
2. García Rodríguez A. Estudio de nuevos marcadores de infertilidad masculina [Tesis de Grado]. Madrid: Universidad Autónoma de Madrid; 2018 [Citado 15/08/2023]. Disponible en: <http://hdl.handle.net/10486/684219>
3. Morey León G, Puga Torres T, Blum Rojas X, González González M, Narváez Sarasti A, Sorroza Rojas N. Caracterización de la calidad del semen en hombres atendidos en un centro de reproducción asistida en Guayaquil, Ecuador. Rev Perú Med Exp Salud Pública [Internet]. 2020 [Citado 15/08/2023];37(2):[Aprox. 4 p.]. Disponible en: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1726-46342020000200292
4. Armas-González E, Mosquera-Escobar M, Alvarez-González K, Rodríguez-Hidalgo M, Duartes-García D, Guerra-Sánchez M. Relationship between risk factors for male infertility and spermogram abnormalities. Rev Ciencias Médicas [Internet]. 2022 Abr [Citado 15/08/2023];26(2):e5500. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-31942022000200021&lng=es.
5. Ramírez Moran AF, Cala Bayeux A, Fajardo Iglesia D, Grave de Peralta RS. Factores causales de infertilidad. Rev Información Científica [Internet]. 2019 [Citado 15/08/2023];98(2):[Aprox. 8 p.]. Disponible en: <http://www.revinfciencia.sld.cu/index.php/ric/article/view/2235>
6. García Valcarce D. Factores que afectan a la calidad espermática y herramientas emergentes en su análisis. [Tesis Doctorado]. León: Universidad de León; 2017 [Citado 15/08/2023]. Disponible en: <https://dialnet.unirioja.es/servlet/tesis?codigo=124342>
7. Amini L, Kahrobaie M, Amiri-Farahani L, Haghani H. The relationship between health life style and spermogram Indicators among infertile men: preliminary data. BMC Res Notes [Internet]. 2020 Jun [Citado 15/08/2023];13(1):278. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32517770>
8. Salas-Huetos A, Rosique-Esteban N, Becerra-Tomás N, Vizmanos B, Bulló M, Salas-Salvadó J. The Effect of Nutrients and Dietary Supplements on Sperm Quality Parameters: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. Adv Nutr [Internet]. 2018 Nov [Citado 15/08/2023];9(6):833–48. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/30462179>
9. Wang C, Swerdloff RS. Limitations of semen analysis as a test of male fertility and anticipated needs from newer tests. Fertil Steril [Internet]. 2014 Dic [Citado 15/08/2023];102(6):1502–7. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/25458617>
10. van der Merwe FH, Kruger TF, Oehninger SC, Lombard CJ. The Use of Semen Parameters to Identify the Subfertile Male in the General Population. Gynecol Obstet Invest [Internet]. 2005 [Citado 15/08/2023];59(2):86–91. Disponible en: <http://dx.doi.org/10.1159/000082368>
11. Rodríguez-Monteaquedo MA, Avello-Martínez R, Avello-Rodríguez A, Reyes-Pérez AM. Análisis de las variables del espermograma en los pacientes de la consulta de infertilidad del municipio de Cienfuegos. Medisur [Internet]. 2021 [Citado 15/08/2023];19(4):[Aprox. 5 p.]. Disponible en: <http://www.medisur.sld.cu/index.php/medisur/article/view/5091>



12. Khandwala YS, Zhang CA, Li S, Cullen MR, Eisenberg ML. Validity of Claims Data for the Identification of Male Infertility. *Curr Urol Rep* [Internet]. 2017 [Citado 15/08/2023];18(9). Disponible en: <http://dx.doi.org/10.1007/s11934-017-0714-7>
13. Espinoza Navarro O, Cortés A S, Monreal J, Ferreccio C. Análisis de las variables del espermiograma en jóvenes sanos en Arica-Chile. *Rev Med Chil* [Internet]. 2010 [Citado 15/08/2023];138(12):1510–6. Disponible en: <http://dx.doi.org/10.4067/s0034-98872010001300006>
14. Macomber D, Sanders MB. The Spermatozoa Count. *N Engl J Med* [Internet]. 1929 [Citado 15/08/2023];200(19):981–4. Disponible en: <http://dx.doi.org/10.1056/nejm192905092001905>
15. World Health Organization Department of Reproductive Health and Research. WHO laboratory manual for the examination and processing of human semen, fifth edition. WHO [Internet]. Geneva: WHO; 2010 [Citado 15/08/2023]. Disponible en: <http://www.who.int/reproductivehealth/publications/infertility/9789241547789/en/>
16. Henao MS, Cardona W. Evaluación de los parámetros seminales en 30 hombres con fertilidad probada y breve revisión de la literatura. *Rev Cubana Obstet Ginecol* [Internet]. 2013 [Citado 15/08/2023];39(4):368-82. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-600X2013000400006&lng=es
17. Wiwie C, Baumbach J, Röttger R. Comparing the performance of biomedical clustering methods. *Nat Methods* [Internet]. 2015 [Citado 15/08/2023];12(11):1033–8. Disponible en: <http://dx.doi.org/10.1038/nmeth.3583>
18. Ibănescu I, Leiding C, Bollwein H. Cluster analysis reveals seasonal variation of sperm subpopulations in extended boar semen. *J Reprod Dev* [Internet]. 2018 Feb [Citado 15/08/2023];64(1):33–9. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29081440>
19. Xu R, Wunsch DC. Clustering Algorithms in Biomedical Research: A Review. *IEEE Rev Biomed Eng* [Internet]. 2010 [Citado 15/08/2023];3:120–54. Disponible en: <http://dx.doi.org/10.1109/rbme.2010.2083647>
20. Meschede D, Lemcke B, Behre HM, De Geyter C, Nieschlag E, Horst J. Clustering of male infertility in the families of couples treated with intracytoplasmic sperm injection. *Hum Reprod* [Internet]. 2000 [Citado 15/08/2023];15(7):1604–8. Disponible en: <http://dx.doi.org/10.1093/humrep/15.7.1604>
21. Meyer H, Reudenbach C, Wöllauer S, Nauss T. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecol Modell* [Internet]. 2019 [Citado 15/08/2023];411:108815. Disponible en: <http://dx.doi.org/10.1016/j.ecolmodel.2019.108815>
22. Avello-Martínez R, Seisdedo-Losa A. El procesamiento estadístico con R en la investigación científica. *Medisur* [Internet]. 2017 [Citado 15/08/2023];15(5):[Aprox. 3 p.]. Disponible en: <http://medisur.sld.cu/index.php/medisur/article/view/3662>
23. Banerjee A, Dave RN. Validating clusters using the Hopkins statistic. En: 2004 IEEE International conference on fuzzy systems; 2004 Jul 25-24 [Internet]. Budapest: PennEstate; 2004 [Citado 15/08/2023];1:149-53. Disponible en: <http://dx.doi.org/10.1109/FUZZY.2004.1375706>



24. Nguyen H, Bui X-N, Tran Q-H, Mai N-L. A new soft computing model for estimating and controlling blast-produced ground vibration based on Hierarchical K-medias clustering and Cubist algorithms. *Appl Soft Comput* [Internet]. 2019 [Citado 15/08/2023];77:376–86. Disponible en: <http://dx.doi.org/10.1016/j.asoc.2019.01.042>
25. Qi J, Yu Y, Wang L, Liu J, Wang Y. An effective and efficient hierarchical K-medias clustering algorithm. *Int J Distrib Sens Networks* [Internet]. 2017 [Citado 15/08/2023];13(8):155014771772862. Disponible en: <http://dx.doi.org/10.1177/1550147717728627>
26. Kimes PK, Liu Y, Neil Hayes D, Marron JS. Statistical significance for hierarchical clustering. *Biometrics* [Internet]. 2017 Sep [Citado 15/08/2023];73(3):811–21. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/28099990>
27. Peterson AD, Ghosh AP, Maitra R. Merging K-medias with hierarchical clustering for identifying general-shaped groups. *Stat (Int Stat Inst)* [Internet]. 2018 [Citado 15/08/2023];7(1):e172. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29736237>
28. Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med* [Internet]. 2019 Feb [Citado 15/08/2023];38(4):558–82. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29869423>
29. Rojas Quintana PR, Guedez Villalobos DD, Reyes Pérez AM, Chávez González N, Rivas Alpizar E. Prevalencia de la chlamydia trachomatis en pacientes sometidas a técnicas de reproducción asistida. *Cienfuegos*, 2016-2018. *Medisur* [Internet]. 2023 Abr [Citado 15/08/2023];21(2):301-8. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1727-897X2023000200301&Ing=es.
30. Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *J Biomed Inform* [Internet]. 2018 [Citado 15/08/2023];85:30–9. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1532046418301308>
31. Patel AS, Leong JY, Ramasamy R. Prediction of male infertility by the World Health Organization laboratory manual for assessment of semen analysis: A systematic review. *Arab Journal of Urology* [Internet]. 2018 [Citado 15/08/2023];16:96–102. Disponible en: <https://doi.org/10.1016/j.aju.2017.10.005>
32. Tejeda-Ruíz C, Panizo-Bruzón S, Blanco-Tejeda A, Sánchez-Cruz Y, Thomas-Santiesteban M. Características de pacientes con espermograma anormal atendidos en la consulta de infertilidad en Las Tunas. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta* [Internet]. 2022 [Citado 15/08/2023];47(6):1-7. Disponible en: <https://revzoilomarinaldo.sld.cu/index.php/zmv/article/view/3236>
33. Frikh M, Benaissa M, Kasouati J, Benlahlou Y, Chokairi O, Barkiyou M, et al. Prevalence of male infertility in a university hospital in Morocco. *Pan Afr Med J* [Internet]. 2021 Ene [Citado 15/08/2023];38:46. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8017356/>
34. Rodríguez Pendás BV, Santana Pérez F, Domínguez Alonso E, Nurquez Guerra B, Reyes Rodríguez H. Leucocitos seminales y calidad espermática de hombres en estudio de infertilidad. *Rev Cubana Endocrinol* [Internet]. 2016 [Citado 15/08/2023];27(1):18-28. Disponible en: <https://scielo.sld.cu/scielo.php?pid=S1561-29532016000100003>



Conflicto de interés

Los autores declaran que no tienen conflictos de intereses.

Declaración de autoría

Mabel A. Rodríguez Monteagudo, Raidell Avello Martínez: conceptualización, curación de datos, metodología, recursos, supervisión.

Mabel A. Rodríguez Monteagudo, Raidell Avello Martínez, Amanda Avello Rodríguez: análisis formal, investigación, administración del proyecto, validación, visualización, redacción – borrador original, redacción – revisión y edición.

Raidell Avello Martínez, Amanda Avello Rodríguez: software.

