



# استفاده از تکنیک داده‌کاوی در بررسی فاکتورهای تأثیرگذار در عدم موفقیت روش انتقال داخل رحمی در درمان ناباروری

فرانک ابوالمعصوم<sup>۱</sup> / سمیه علیزاده<sup>۲</sup> / محسن اصغری<sup>۳</sup>

## چکیده

**مقدمه:** ناباروری یکی از مسائلی است که امروزه هزینه‌های مادی و معنوی زیادی را بر روی دوش زوجین نابارور قرار داده است. روش انتقال داخل رحمی اسپرم (Intra-Uterine Insemination (IUI)) یکی از روش‌های درمانی برای این گروه از زوجین می‌باشد. غیرقابل پیش‌بینی بودن نتیجه این روش ضرورت بررسی و شناسایی عواملی را که بر روی میزان اثربخشی آن تأثیرگذار هستند، دوچندان کرده است. لذا، هدف این مطالعه شناسایی عوامل مؤثر در عدم موفقیت این روش با بکارگیری تکنیک‌های داده‌کاوی روی داده‌های مربوط به زوجین نابارور بیمارستان صارم تهران می‌باشد که از روش IUI استفاده کرده و نتیجه این روش بر روی آن‌ها ناموفق بوده است.

**روش کار:** بیماران به هفت گروه تقسیم شدند و با استفاده از تکنیک توصیفی داده‌کاوی یعنی الگوریتم K means و شاخص Davis-Buldian ویژگی‌های هر گروه مورد تحلیل و بررسی قرار گرفت.

**یافته‌ها:** بررسی و تحلیل ویژگی‌های هفت گروه بدست آمده، نشان داد که: افزایش سن خانم، اضافه وزن و یا چاقی خانم دو عامل اصلی عدم موفقیت روش درمانی IUI بودند. عامل بعدی، نوع ناباوری است، که در این مطالعه، اکثریت زوجین، نوع ناباروری اولیه را تجربه کرده بودند. افزون بر این، فاکتور مردانه فاکتور دیگری بود که روی عدم موفقیت روش IUI تأثیر داشته است. نتایج نشان داد که طول مدت ناباروری، و وضعیت اسپرم از لحاظ نرمال یا غیر نرمال بودن، از دیگر عوامل شکست روش درمانی مذکور هستند.

**نتیجه‌گیری:** در این پژوهش با تحلیل نتایج خوشه‌بندی، عوامل تأثیرگذار در عدم موفقیت روش IUI بر روی زوجین نابارور شناسایی شد. این نتایج همراه با مشورت با خبرگان می‌تواند برای پیش‌بینی نتیجه روش IUI و انتخاب روش درمانی مناسب به محققین، پزشکان و زوجین نابارور کمک کند.

**کلید واژه‌ها:** ناباروری، روش IUI، داده‌کاوی، روش‌های توصیفی، الگوریتم Kmeans

• وصول مقاله: ۹۲/۴/۲۹ • اصلاح نهایی: ۹۲/۸/۱۳ • پذیرش نهایی: ۹۲/۸/۲۹

۱. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران، نویسنده مسئول (faranakabolmasum@yahoo.com)

۲. استادیار گروه مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

۳. کارشناسی ارشد، مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

بر مورد (Case Based Reasoning (CBR)) را به منظور پیشنهاد اصلاحات امکان پذیر در پروسه درمانی IVF بر اساس تجارب گذشته ارائه کردند تا نرخ نهایی موفقیت در روش IVF بهبود یابد [۸]. در سال‌های اخیر داده‌های کاوی در حوزه ناباروری با استقبال بیشتری از سوی محققان داخلی مواجه شده است. از جمله این مطالعات می‌توان به طراحی یک سیستم خبره برای پیش‌بینی نتیجه حاصل از سیکل درمانی IUI [۴] و نیز ارائه یک مدل تشخیصی بر اساس استنتاج مبتنی بر مورد (CBR) با استفاده از داده‌های مربوط به زوجین نابارور [۹] اشاره کرد. با توجه به مطالعات انجام شده دیده می‌شود که تمرکز بر روی روش درمانی IUI کمتر بوده و بررسی فاکتورهای اثرگذار در شکست آن به ندرت مورد توجه محققین قرار گرفته است. با توجه به اینکه روش‌های درمان ناباروری هزینه‌های مادی و معنوی زیادی را بر دوش زوجین نابارور تحمیل می‌کند، با شناسایی فاکتورهای اثرگذار در عدم موفقیت این روش‌ها به ویژه روش IUI که یکی از روش‌های کمک باروری Assisted Reproductively Technology (ART) می‌باشد می‌توان این هزینه‌ها را به شکل چشمگیری کاهش داد و در انتخاب روش مناسب درمانی به پزشکان، محققین و زوجین نابارور کمک کرد. از این رو در این مطالعه روی این روش تمرکز شد و هدف بررسی داده‌های مربوط به زوجین نابارور بیمارستان صارم می‌بود که از روش IUI استفاده کرده و نتیجه این روش روی آن‌ها ناموفق بوده است. با بررسی ویژگی‌های این افراد با استفاده از تکنیک‌های توصیفی داده کاوی، می‌توان برخی فاکتورهای تأثیرگذار در عدم موفقیت روش IUI را شناسایی کرد. با شناسایی این فاکتورها می‌توان به مطالعات آتی در زمینه پیش‌بینی میزان موفقیت روش IUI و در صورت امکان رفع فاکتورهای اثرگذار در شکست این روش و نیز انتخاب روش درمانی مناسب به محققان، پزشکان و همچنین زوجین نابارور کمک کرد.

### روش کار

ماهیت مطالعه حاضر داده‌محور می‌باشد و پایه اصلی آن بر داده کاوی و تحلیل و بررسی روی داده‌های ناباروری موجود

### مقدمه

امروزه با توجه به گسترش کاربرد فناوری اطلاعات در حوزه‌های مختلف و در نتیجه افزایش سریع حجم داده‌ها، به کارگیری ابزاری برای کشف و استخراج الگوهای بین آن‌ها بسیار حائز اهمیت است. داده کاوی تکنیکی است که می‌تواند در این زمینه اثربخش باشد. داده کاوی به عنوان بخشی از فرایند کشف دانش، شامل این مراحل است: شناسایی محدوده کسب و کار، انتخاب داده‌ها، پاک سازی داده‌ها، کشف الگوها که شامل استخراج الگوهای دانش است و در نهایت تفسیر و نمایش الگوها. هم چنین روش‌های اصلی داده کاوی به دو دسته الگوریتم‌های پیش‌بینانه (نظارتی) و توصیفی (غیرنظارتی) تقسیم می‌شوند [۱]. روش‌ها و الگوریتم‌های توصیفی به منظور یافتن الگوهای قابل تفسیر برای انسان در مورد داده‌ها استفاده می‌شود و در این مطالعه از این روش استفاده شده است.

حوزه ناباروری در پزشکی از جمله حوزه‌هایی است که می‌توان با به کارگیری داده کاوی در آن به نتایج بسیار مفیدی دست یافت [۲]. طبق آمارهای رسمی منتشر شده ۹/۲ درصد از زوج‌های ایرانی ناباروری اولیه را در طول زندگی مشترک خود تجربه می‌کنند [۳،۴]. این امر در حالی است که طبق گزارش سازمان جهانی بهداشت ((World Health Organization))، میزان ناباروری در کشورهای مختلف بین پنج تا ۳۰ درصد می‌باشد. در نتیجه می‌توان گفت ایران جزو کشورهایی با ریسک نسبتاً بالا در این زمینه محسوب می‌شود. لذا، مطالعه در زمینه ناباروری در ایران از اهمیت بالایی برخوردار است. تاکنون مطالعات متعددی در حوزه کاربرد تکنیک‌های داده کاوی در زمینه ناباروری در خارج و داخل ایران صورت گرفته است که می‌توان به ارائه مدل‌های ریاضی یا سیستم‌های هوشمند برای بهبود نرخ پیش‌بینی نتایج روش درمانی In Vitro Fertilization (IVF) [۵،۶] و یا کاربرد تکنیک‌های هوش مصنوعی برای پیش‌بینی نتایج همین روش درمانی [۷] اشاره کرد. مطالعه انجام شده توسط ایگور جوریسیکا نیز مثال دیگری است که یک سیستم استنتاج مبتنی

در پایگاه داده‌های بیمارستان صارم تهران که از سال ۱۳۷۶ تا سال ۱۳۸۸ جمع آوری شده، بنا نهاده شده‌است. لذا، مدل CRISP-DM جهت انجام فرایند تحقیق مورد استفاده قرار گرفته‌است. در این بخش فرایند اجرایی تحقیق بر اساس مراحل این مدل شرح داده می‌شود.

#### شناخت داده‌ها

بیماران به منظور درمان‌های مختلفی از جمله IUI و IVF به بیمارستان صارم مراجعه می‌نمایند و اطلاعات شخصی زوجین و نتایج آزمایشات در پرونده‌های آن‌ها ثبت و نگهداری می‌شود [۹].

#### آماده سازی داده‌ها

این گام از فرایند تحلیل، با عنوان پیش پردازش داده‌ها شناخته می‌شود. پیش پردازش داده‌ها جهت بهبود کیفیت داده‌های واقعی برای داده‌کاوی لازم است [۱۰]. ۷۵ تا ۹۰ درصد موفقیت پروژها به مرحله آماده سازی و پیش پردازش داده‌ها وابسته است [۱۱]. این مرحله شامل انتخاب منابع و پایگاه‌های داده، حذف نقاط پرت یا مغشوش، طرز برخورد با داده‌های گمشده (برای مثال حذف نمونه‌هایی که مقادیر گمشده دارند، یا پر کردن مقادیر گمشده به طور دستی و روش‌های مختلف دیگر که با توجه به نوع و حجم پایگاه داده می‌توان یکی از آن‌ها را برگزید)، تبدیل داده (به این معنا که داده‌ها را به شکلی که برای انجام داده‌کاوی مناسب است، تبدیل کرد)؛ و کاهش داده‌هاست. باید گفت تکنیک‌های کاهش داده برای دستیابی به یک کاهش در داده‌ها به کار می‌روند، به طوریکه داده‌های جدید علاوه بر کم حجم‌تر بودن نسبت به داده‌های اولیه، به همان میزان یکپارچه باشند. داده با کیفیت باید صحیح، کامل، سازگار، به‌روز، قابل قبول، باارزش، قابل تفسیر و در دسترس باشد [۱۰]. در این مطالعه داده‌های مربوط به بیمارانی مورد بررسی قرار گرفت که در بیمارستان صارم عمل IUI بر روی آن‌ها انجام شده بود، از بین فیله‌های موجود در پایگاه داده، برخی فیله‌های مؤثرتر که ۱۳ عدد می‌باشند، براساس مطالعات گذشته در این زمینه [۴، ۹] و نیز مشورت با خبرگان این رشته به دست آمده، انتخاب و به کار گرفته شد. در ادامه در مورد این ۱۳ ویژگی توضیحاتی ارائه شده است.

اولین ویژگی سن خانم است که مطالعات نشان می‌دهد که با بالا رفتن سن خانم‌ها، احتمال موفقیت سیکل IUI کاهش می‌یابد [۴]. عوامل زنانه ویژگی بعدی است. عوامل زنانه در ناباروری را می‌توان به پنج دسته تقسیم بندی نمود: اختلال در تخمک گذاری، اختلال در لوله‌های رحمی، آندومتریوز، اختلال در گردن رحم و مشکلات ایمنی بدن. ویژگی بعدی عامل مردانه است؛ عوامل ناباروری در مردان نیز می‌تواند به دلایلی گوناگون از جمله اختلالات مربوط به اسپرم، بیماری‌های عفونی، اختلال هورمونی و مشکلات مربوط به دستگاه ایمنی بدن باشد. میانگین تعداد فولیکول‌های سمت چپ و راست نیز ویژگی مؤثر دیگری است که انتخاب شده است، این تعداد مربوط به روزی است که آمپول BHCG جهت تخمک گذاری، در فرایند IUI تزریق می‌شود. ویژگی منتخب بعدی نوع ناباروری است، ناباروری به دو نوع اولیه (بدون سابقه بارداری در گذشته) و ثانویه (با سابقه بارداری در گذشته) تقسیم می‌شود. طول مدت ناباروری، شاخص توده بدن خانم (Body Mass Index (BMI که از تقسیم وزن (بر حسب کیلوگرم) بر مجذور قد (بر حسب متر) به دست می‌آید. ویژگی ضخامت آندومتر نیز برای این کار انتخاب شد. این ویژگی هنگام انجام عمل IUI به نحوه لانه‌گزینی جنین اشاره داشته و روی میزان موفقیت و عدم موفقیت این روش تأثیرگذار است. تعداد اسپرم‌ها بعد از شستشو نیز جزو ویژگی‌های منتخب است. تعداد اسپرم‌ها بعد از شستشوی آزمایشگاهی، میزان اسپرم تزریقی به خانم را مشخص می‌کند. ویژگی تعداد اسپرم‌های نرمال بعد از شستشو بیانگر درصدی از اسپرم‌هاست که شکل ظاهری نرمال دارند و با موفقیت سیکل IUI رابطه مستقیم دارد. تعداد اسپرم‌های Full که از لحاظ حرکتی کاملاً طبیعی هستند و Sluggish دارای حرکت دورانی بوده و مستقیم حرکت نمی‌کند [۱۲]؛ و آخرین ویژگی منتخب نیز وضعیت نرمال یا غیر نرمال بودن اسپرم را مشخص می‌کند.

در نهایت، بعد از آماده سازی داده‌ها و پاک سازی داده‌ها، مانند حذف پرونده‌هایی که سن آن‌ها از طیف معمولی خود خارج بودند که این خطا در هنگام ورود اطلاعات توسط

مشخص نیست یافتن تعداد خوشه‌ی بهینه یکی از مسائلی موجود برای خوشه بندی داده‌های منتخب بود. برای یافتن تعداد  $K$  بهینه الگوریتم  $K$  means با ورودی  $K=2$  تا  $K=8$  اجرا گردید و سپس برای تعیین تعداد خوشه بهینه از شاخص دیویس - بولدین (Davies - Bouldin) که شاخصی معتبر در این زمینه است، استفاده شد. این شاخص تابعی از نسبت جمع پراکندگی داخل خوشه به جدایی خوشه‌ها از هم می‌باشد [۲،۹،۱۰]. در معادله (۱) فرض می‌شود که  $v$  مرکز خوشه،  $q$  یک عدد صحیح و  $|A_i|$  تعداد عناصر  $A_i$  است.

$$S_{i,q} = \left( \frac{1}{|A_i|} \sum_{x \in A_i} \|x - v_i\|_2^q \right)^{1/q} \quad \text{معادله (۱)}$$

$$d_{ij,s} = \left\{ \sum_{s=1}^p |v_{si} - v_{sj}|^t \right\}^{\frac{1}{t}} = \|v_i - v_j\|_t \quad \text{معادله (۲)}$$

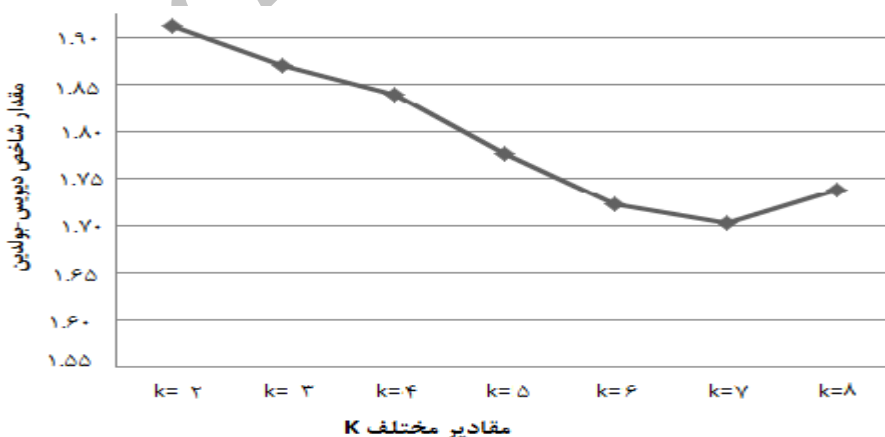
و

$$R_{i,qt} = \max_{j \in c, j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{d_{ij,s}} \right\} \quad \text{معادله (۳)}$$

و در نهایت شاخص دیویس - بولدین به صورت زیر تعریف می‌شود:

$$DB(c) = \frac{1}{2} \sum_{j=1}^c R_{j,qt} \quad \text{معادله (۴)}$$

هرچه مقدار این شاخص برای تعداد خوشه متناظرش کم باشد، آن تعداد خوشه بهینه تر می‌باشد. (نمودار ۱):



نمودار ۱: مقدار شاخص دیویس - بولدین به ازای  $K=2$  تا  $K=8$

کاربر رخ داده است، از بین پرونده‌های موجود تعداد ۳۳۹ پرونده از سال‌های ۱۳۷۶ تا ۱۳۸۸ در بیمارستان صارم انتخاب شدند. که نتایج روش IUI بر روی آن‌ها ناموفق بوده است این پرونده‌ها دارای داده‌های مناسب بوده و به تأیید خبرگان برای انجام تحقیقات رسیده است. منظور از عدم موفقیت در این روش منفی بودن تست بتا و عدم مشاهده ساک حاملگی در بیمار می‌باشد.

#### مدل سازی داده‌ها

در این گام، روشی برای مدل سازی داده‌های منتخب و پاک سازی شده انتخاب شد. از آنجایی که تمرکز این مطالعه بر روی روش‌های توصیفی (غیر نظارتی) می‌باشد، از یکی از رایج ترین الگوریتم‌های توصیفی یعنی الگوریتم خوشه بندی برای تحلیل داده‌های ناباروری و کشف الگوهای مشابه بین دسته‌های مختلف استفاده شده است. هدف اساسی در این مدل سازی رسیدن به یک تحلیل کلی از داده‌ها و نوعی تقسیم بندی بر روی آن‌ها بوده است. به همین منظور، از الگوریتم  $K$  means برای خوشه بندی استفاده شده است. در این الگوریتم داده‌ها به  $K$  گروه مختلف توزیع می‌شوند که داده‌های هر گروه با یکدیگر مشابه بوده و داده‌های گروه‌های مختلف با یکدیگر نامتشابه است. الگوریتم  $K$  means تعداد خوشه‌ها را از ورودی دریافت می‌کند و از آنجایی که تعداد خوشه بهینه

با توجه به نتیجه این شاخص و نظر خبرگان در این زمینه، و بر اساس خوشه‌بندی با الگوریتم  $K$  means داده‌های مربوط به

همان طور که در نمودار (۱) نیز مشخص است تعداد خوشه - های بهینه بر اساس شاخص دیویس - بولدین، هفت می‌باشد.

**یافته‌ها**

بیماران در هر یک از گروه‌ها دارای ویژگی‌های مشابهی می‌باشند. جدول (۳) ویژگی‌های هر یک از این گروه‌ها را نمایش می‌دهد و در ادامه به بررسی ویژگی‌های هر یک از گروه‌ها پرداخته می‌شود.

بیماران در هفت گروه تقسیم‌بندی شدند. با بررسی نتایج به کارگیری الگوریتم K means پراکندگی بیماران در هر گروه بدین ترتیب می‌بود: بیماران در گروه اول ۷۹ نفر، در گروه دوم ۲۱ نفر، در گروه سوم ۲۷ نفر، در گروه چهارم ۲۸ نفر، در گروه پنجم ۶۹ نفر، در گروه ششم ۵۹ نفر و در گروه هفتم ۴۶ نفر.

**جدول ۱: ویژگی‌های مربوط به هفت گروه منتخب بیمار**

ویژگی	گروه اول	گروه دوم	گروه سوم	گروه چهارم	گروه پنجم	گروه ششم	گروه هفتم
میانگین سن (سال)	۲۸/۶	۲۷/۱	۲۵/۳	۲۶/۶	۳۱/۹	۲۸/۶	۲۹/۵
متوسط مدت ناباروری (سال)	۴/۶	۴	۳/۹	۳/۵	۳/۴	۳/۶	۵
نوع ناباروری غالب	۱۰۰٪ اولیه	۵۷/۲٪ اولیه	۱۱/۲٪ اولیه	۱۰۰٪ اولیه	۱۰۰٪ ثانویه	۱۰۰٪ اولیه	۳۷٪ اولیه ۶۳٪ ثانویه
BMI < ۱۸/۵	-	-	-	۱۰/۶٪	-	-	-
وضعیت BMI	-	۱۹/۱٪	۵۹/۳٪	۸۹/۵٪	۳۷/۷٪	۱۰۰٪	۱۳/۱٪
۱۸/۵ ≤ BMI < ۲۴/۹	-	۳۸/۱٪	۱۸/۶٪	-	۴۳/۵٪	-	۳۰/۵٪
۲۵ ≤ BMI < ۲۹/۹	۱۲/۷٪	۸۶/۹٪	۲۲/۲۲٪	-	۱۸/۹٪	-	۵۶/۶٪
۳۰ ≤ BMI	۹۸/۸٪	-	۱۰۰٪	۸۶/۹٪	۹۸/۶٪	۹۱/۶٪	۱۰۰٪
علت ناباروری	MF	OF	OF	OF	OF	OF	OF
نامشخص	۱/۳٪	۱۴/۳٪	-	-	۱/۵٪	۳/۴٪	-
جمع تعداد فولیکول‌های سمت راست و چپ هنگام عمل	میانگین ۳/۱ عدد	میانگین ۴/۳ عدد	میانگین ۵/۱ عدد	میانگین ۵/۳ عدد	میانگین ۳/۴ عدد	میانگین ۳/۷ عدد	میانگین ۴/۵ عدد
میانگین ضخامت آندومتر هنگام عمل	۸/۲ میلی‌متر	۸ میلی‌متر	۸/۲ میلی‌متر	۷/۶ میلی‌متر	۷/۵ میلی‌متر	۷/۷ میلی‌متر	۷/۷ میلی‌متر
تعداد اسپرم‌های نرمال بعد از شستشو	میانگین ۴/۱ میلیون	میانگین ۵/۴ میلیون	میانگین ۲/۷ میلیون	میانگین ۵/۳ میلیون	میانگین ۴/۲ میلیون	میانگین ۳/۸ میلیون	میانگین ۴/۷ میلیون
تعداد اسپرم‌های Full بعد از شستشو	میانگین ۳/۶ میلیون	میانگین ۵/۵ میلیون	میانگین ۲ میلیون	میانگین ۴/۷ میلیون	میانگین ۳/۸ میلیون	میانگین ۳/۲ میلیون	میانگین ۳/۹ میلیون
تعداد اسپرم‌های Sluggish بعد از شستشو	میانگین ۲/۸ میلیون	میانگین ۲/۸ میلیون	میانگین ۲/۵ میلیون	میانگین ۳/۳ میلیون	میانگین ۲/۸ میلیون	میانگین ۲/۷ میلیون	میانگین ۳/۲ میلیون
نوع غالب Spermogram type	۷۴/۷ درصد Terato Spermia	۷۱/۵ درصد Spermogram Norma	۸۵/۲ درصد Pyo Spermia	۹۷/۴ درصد Terato Spermia	۷۱ درصد Terato Spermia	۷۹/۷ درصد Terato Spermia	۷۹/۷ درصد Terato Spermia

اضافه وزن بوده و تمامی آن‌ها ناباروری اولیه را تجربه کرده‌اند و بیش از ۹۸ درصد از زوجین در گروه اول بیماران (به دلیل فاکتور مردانه) دچار ناباروری بودند و ۳۵/۵ درصد از

گروه اول بیماران: این گروه بیشترین فراوانی از نظر تعداد را در بین سایر گروه‌ها دارد. همان‌طور که در جدول سه نیز مشاهده می‌شود، بیشترین درصد افراد در این گروه دارای

گروه چهارم بیماران: کمترین میانگین سنی را در بین سایر گروه‌ها را داشت. درصد افراد دارای وزن طبیعی درصد بالایی را در این گروه به خود اختصاص داده است. تمامی زوجین در این گروه از بیماران، ناباروری اولیه را تجربه کرده و به دلیل تخمک گذاری نابارور بودند. ۸۶/۸ درصد از آن‌ها نیز علاوه بر عامل تخمک گذاری دارای عامل مردانه نیز بودند. وضعیت اسپرم این گروه در جدول (۲) نمایش داده شده است. همان طور که در جدول نیز مشخص است بیشترین درصد افراد در این گروه وضعیت اسپرم Terato Spermia داشته‌اند.

گروه پنجم بیماران: بیشترین میانگین سنی و کمترین طول مدت ناباروری را در بین سایر گروه‌ها داشتند. اکثریت افراد در این گروه دارای اضافه وزن و چاقی بودند و تمامی آن‌ها نوع ناباروری ثانویه را تجربه کرده بودند. ۹۸/۶ درصد از آن‌ها به دلیل فاکتور مردانه و ۱/۵ درصد باقی مانده به دلیل نامشخص نابارور بودند. این گروه کمترین ضخامت آندومتر را در بین سایر گروه‌ها داشتند. وضعیت اسپرم این گروه از بیماران در جدول (۲) نشان داده شده است. همان‌طور که در جدول نیز مشخص است، بیشترین درصد وضعیت اسپرم Terato Spermia است.

گروه ششم بیماران: تمامی افراد دارای وزن طبیعی بوده و نوع ناباروری اولیه را تجربه کرده بودند. ۹۶/۶۱ درصد از زوجین در این گروه به دلیل فاکتور مردانه و ۳/۴ درصد نیز به دلیل نامشخص نابارور هستند. وضعیت اسپرم در این گروه نیز در جدول (۲) نشان داده شده است. همان‌طور که مشخص است بیشترین درصد وضعیت اسپرم متعلق به وضعیت Terato Spermia بود.

گروه هفتم بیماران: این گروه از بیماران دارای بیشترین مدت زمان بارداری در بین سایر گروه‌ها بودند. اکثریت افراد در این گروه اضافه وزن و چاقی داشتند. ۶۳ درصد از آن‌ها ناباروری ثانویه را تجربه کرده بودند. تمامی زوجین در این گروه هم به دلیل فاکتور مردانه و هم تخمک گذاری نابارور بوده‌اند. بیشترین وضعیت اسپرم در این گروه، همان‌طور که در جدول (۲) نیز نشان داده شده است، Terato Spermia است.

آن‌ها علاوه بر فاکتور مردانه عامل تخمک گذاری نیز دارند. ۱/۳ میانگین جمع تعداد فولیکول‌های سمت راست و چپ هنگام عمل در این گروه در مقایسه با سایر گروه‌ها کمترین میزان را داشتند. وضعیت اسپرم در این گروه از بیماران همان‌طور که در جدول دو نیز مشخص است به طوری است که تنها کمتر از دو درصد افراد در این گروه وضعیت اسپرم نرمال داشته و وضعیت اسپرم سایر افراد گروه غیر نرمال بود و بیشترین مشکل اسپرم در افراد این گروه مشکل Terato Spermia بود. این مشکل اشاره به وجود مشکل در شکل ظاهری اسپرم دارد [۱۲].

گروه دوم بیماران: نیز همان‌طور که در جدول (۳) نیز مشخص است، نزدیک به نیمی از افراد در این گروه از بیماران (۴۲/۹ درصد) چاق بوده و بیش از نیمی از آن‌ها ناباروری اولیه را تجربه کرده بودند. ۸۵/۷ درصد از زوجین این گروه بیمار، به دلیل عامل تخمک گذاری نابارور بودند. این گروه از بیماران بیشترین تعداد اسپرم‌های نرمال از لحاظ شکل و ظاهر و نیز اسپرم‌های Full بعد از شستشو را در بین سایر گروه‌ها به خود اختصاص داده بودند. وضعیت اسپرم از لحاظ نرمال یا غیرنرمال بودن در این گروه بیمار، در جدول دو نمایش داده شده است. همان‌طور که در جدول نیز مشخص است بیشتر بیماران در این گروه وضعیت اسپرم نرمال داشته‌اند.

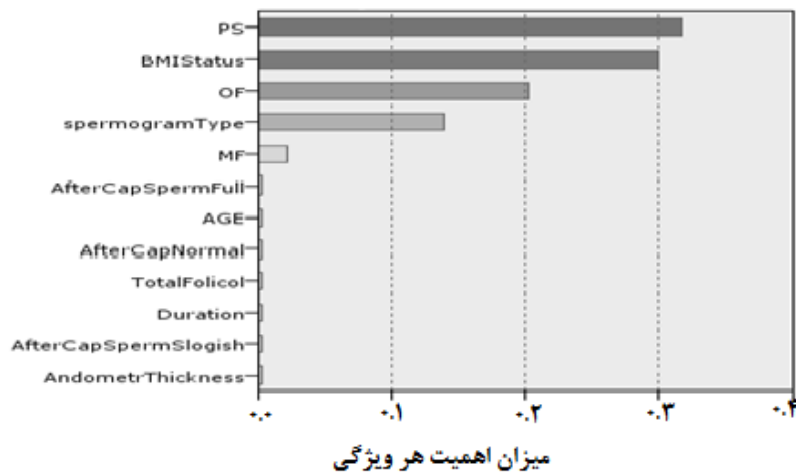
گروه سوم بیماران: نیز همان‌طور که در جدول سه مشخص است، بیش از نیمی از افراد در این گروه دارای وزن طبیعی بودند و نیز درصد بالایی از زوجین در این گروه، ناباروری ثانویه را تجربه کرده بودند. تمامی زوجین در این گروه دارای ناباروری با عامل مردانه و تمامی آن‌ها علاوه بر فاکتور مردانه عامل تخمک گذاری نیز داشته‌اند. ضخامت آندومتر در این گروه در مقایسه با سایر گروه‌ها بیشترین میزان را داشت. اما تعداد اسپرم‌ها (نرمال، Full و Sluggish) در این گروه کمتر از سایر گروه‌ها بود. وضعیت اسپرم در این گروه در جدول (۲) نمایش داده شده است. همان‌طور که در جدول نیز مشخص است، ۸۵ درصد از افراد گروه وضعیت Pyo Spermia داشتند که اشاره به مشکل میکروبی در اسپرم دارد [۱۲].

جدول ۲: وضعیت Spermogram Type در هفت گروه بیمار

نام گروه	وضعیت Spermogram Norma	وضعیت Astheno Spermia	وضعیت Terato Spermia	وضعیت Astereto Spermia	وضعیت Pyo Spermia
گروه اول	۱/۲۷ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	۷۴/۷ درصد	۱/۳ درصد	۲۲/۸ درصد
گروه دوم	۷۱/۴ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	۲۸/۶ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	این وضعیت در این گروه نمونه‌ای نداشته است
گروه سوم	این وضعیت در این گروه نمونه‌ای نداشته است	این وضعیت در این گروه نمونه‌ای نداشته است	این وضعیت در این گروه نمونه‌ای نداشته است	۱۵ درصد	۸۵ درصد
گروه چهارم	۲/۶ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	۹۷/۴ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	این وضعیت در این گروه نمونه‌ای نداشته است
گروه پنجم	این وضعیت در این گروه نمونه‌ای نداشته است	۳/۷ درصد	۷۱ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	۲۱/۸ درصد
گروه ششم	این وضعیت در این گروه نمونه‌ای نداشته است	۳/۴ درصد	۷۹/۷ درصد	این وضعیت در این گروه نمونه‌ای نداشته است	۱۷ درصد
گروه هفتم	این وضعیت در این گروه نمونه‌ای نداشته است	این وضعیت در این گروه نمونه‌ای نداشته است	۹۳/۵ درصد	۲ درصد	۴/۴ درصد

درمانی روی گروه بیماران دارند. نمودار (۲) این ویژگی‌ها و ترتیب اهمیت آن‌ها روی گروه بندی بیماران و نیز تأثیر روی عدم موفقیت روش IUI را نشان می‌دهد.

با توجه به شرح ویژگی‌های هر یک از هفت گروه، باید به این نکته اشاره کرد که برخی ویژگی‌ها تأثیر بیشتری در تفاوت هر یک از گروه‌ها از یکدیگر و نیز در عدم موفقیت روش



نمودار ۲: ترتیب اهمیت ویژگی‌ها در گروه بندی بیماران

تعداد کم فولیکول و حرکت ضعیف اسپرم به عنوان عوامل تأثیرگذار در عدم موفقیت عمل IUI اشاره شده است. در نتیجه تحلیل این ویژگی‌ها به موارد زیر حاصل شد: نوع ناباروری غالب که درصد بالایی از این بیماران بارداری اولیه داشته، اضافه وزن و چاقی با توجه به اینکه درصد بالایی از بیماران در این بررسی دارای اضافه وزن و چاقی بودند، علت ناباروری که درصد بالایی از این بیماران به دلیل عامل مردانه نابارور بوده‌اند، وضعیت غیرنرمال اسپرم که در این گروه از بیماران بیشترین مشکل اسپرم مشکل Terato Spermia بوده‌است و طول مدت ناباروری، روی عدم موفقیت عمل IUI در این بیماران تأثیر داشته‌است. البته ذکر این نکته ضروریست که مطالعه انجام شده محدود به داده‌های دردسترس مربوط به زوجین نابارور بیمارستان صارم می‌باشد که از روش IUI استفاده کرده‌اند، بنابراین حوزه تحقیق و نتایج به دست آمده در محدوده همین داده‌ها می‌باشد. لذا، می‌توان برای مطالعات آتی حوزه تحقیق را گسترده‌تر کرده و داده‌های بیماران مختلف از مراکز مختلف را مورد بررسی قرار داد. همچنین ویژگی‌های روحی و تأثیرات محیطی نیز روی نتایج عمل IUI تأثیرگذار است که به عنوان یکی از موضوعات پیشنهادی برای مطالعات آتی می‌باشد. از نتایج این تحقیق می‌توان در آینده برای پیش‌بینی نتیجه عمل IUI استفاده و به محققان و پزشکان در انتخاب روش درمانی مناسب کمک کرده و نیز هزینه‌های مادی و معنوی وارده به زوجین نابارور را کاهش داد.

### تشکر و قدردانی

از زحمات جناب آقای دکتر صارمی ریاست محترم بیمارستان فوق تخصصی زنان صارم و جناب آقای دکتر صالحیان و همچنین کادر متخصص و با کفایت آن بیمارستان که در مراحل مختلف این پروژه ما را یاری نمودند، و سرکار خانم شامی کمال تشکر و قدردانی را داریم.

همانطور که در شکل نیز مشخص است، به ترتیب نوع ناباروری (PS)، فاکتور حجم توده بدن، ناباروری با علت زنانه (OF) و نوع Spermogram بیشترین تأثیر را در گروه بندی بیماران و عدم موفقیت IUI در این مطالعه داشته‌اند. سایر ویژگی‌ها نیز تأثیرگذار بوده ولی میزان اثرگذاری آن‌ها به مراتب کمتر از چهار ویژگی مذکور، است.

### بحث و نتیجه گیری

هریک از گروه بیماران به دست آمده دارای ویژگی‌هایی است که می‌تواند بر روی نتیجه عمل IUI تأثیرگذار باشد. به طور کلی گروه سنی خانم‌ها در تمامی گروه بیماران بین ۲۷ تا ۳۱ سال است. در مطالعه مشابه انجام شده [۱۲] نیز میانگین سنی زنان همین میزان بوده است طول مدت ناباروری زوجین در بازه زمانی سه تا پنج سال بود. با دقت در وضعیت BMI افراد در این گروه‌ها این نتیجه به دست آمد که اکثریت بیماران دارای اضافه وزن و یا چاقی هستند. به طور کل نوع ناباروری غالب در بیشتر گروه‌ها ناباروری اولیه و عامل ناباروری در آن‌ها فاکتور مردانه بود که در مطالعات مشابه [۱۳، ۱۴] نیز عامل مردانه بیشترین دلیل ناباروری زوجین بوده است. با بررسی وضعیت اسپرم از نظر نرمال یا غیر نرمال بودن آن، باید گفت که تنها در صورتی که عامل ناباروری تخمک گذاری بوده باشد، وضعیت اسپرم نرمال است و در غیر این صورت وضعیت اسپرم غیرنرمال بوده و بیشترین درصد مشکل اسپرم را در بین تمامی گروه‌ها، وضعیت Terato Spermia و بعد از آن Pyo Spermia داشته است. با توجه به این نتایج و نیز مطالعات انجام شده در زمینه عوامل تأثیرگذار در میزان موفقیت عمل IUI، می‌توان نتیجه گرفت که عواملی هم‌چون نوع ناباروری، اضافه وزن و چاقی، عامل ناباروری، وضعیت غیرنرمال اسپرم، سن که در مطالعه انجام شده توسط هیوز [۱۵] نیز به آن اشاره شده و طول مدت ناباروری اشاره شده، در مطالعه پی لاسکر و دیگران [۱۶] تأثیر بیشتری بر روی نتیجه عمل IUI روی این بیماران داشته است. در سایر مطالعات از جمله مطالعه تام لیسون [۱۷] به کاهش ضخامت آندومتر،



## References

- Han J, Kamber M, Kaufmann M, 2nd ed. Data Mining: Concepts and Techniques. 2006.
- Hoseini M. [Developing a predictive model based on the Sarem hospital infertility data][MSc thesis]. Tehran: K.N. Toosi University of technology; 2012. [Persian]
- Vahidi S, Ardalan A, Mohammad K. [Prevalence of primary infertility in Iran in 1383 to 84]. Faslname-Ye Barvari va nabarvari; 2006; pp, 243-251. [Persian]
- Rahbar M. [Design of an expert system to predict the outcome of infertility treatment cycle] [MSc thesis]. Tehran: Tabatabayie University; 2010.[Persian]
- Trimarchi J.R, Goodside J, Passmore L, Silberstein T, Hamel L, Gonzalez L. Comparing Data Mining and Logistic Regression for predicting IVF Outcome. 59<sup>th</sup> Annual meeting of the American society for Reproductive Medicine (ASRM); 2003
- Uyar A, Nadir Ciray H, Bener A, Bahceci M. 3P: personalized Pregnancy Prediction in IVF Treatment Process. eHealth, LNICST 1, 2009; p. 58-65
- Girela J.L, Gil D, Johnsson M, Gomez-Torres M.J, Juan J.D. Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence. Biology of Reproduction; 2013; Vol.88
- Igor Jurisica A, John Mylopoulos A, Glasgow J, Heather Shapiro C, Robert F. Casper. Case-based reasoning in IVF: prediction and knowledge mining. Artificial Intelligence in Medicine; 1998; 12: 1-24.
- Asghari M. [Providing a diagnostic model based on CBR using Sarem infertility data] [MSc thesis]. Tehran: K.N. Toosi University of technology; 2012. [Persian]
- Aghabeigi N. [Providing a hybrid model for clustering on infertility data] [MSc thesis]. Tehran: K.N. Toosi University of technology; 2012. [Persian]
- Plye D. Chapter 14: Data collection, Preparation, Quality, and Visualization. The Handbook of Data Mining, Edited by Ye N., Lawrence Erlbaun Associates, Inc; 2003
- Saremi A. [Introduction to Infertility]. Tehran; 2004. [Persian]
- Sedghiani M, Ayatollahi H, Nanbakhsh F, Mojahediyeh A. [the success rate of IUI treatment in IVF center of Motahari Hospital, 1380 ]. Journal of Ardebil Uuniversity of Medical science; 2006; 2:142-148. [Persian]
- Yousefi Z, Torabi zadeh A, Khadem M, Jahanian M. [The prevalence and success of IUI treatment in infertile couples with different causes of infertility in the center of Mashhad University of medical sciences]. Journal of Mashhad University of Medical science; 2001; 71:19-24. [Persian]
- Hughes EG. The effectiveness of ovulation induction and intrauterine insemination in the Treatment of persistent infertility: a meta-analysis. Hum Reprod. 1997 Sep; 12 (9): 1865-72.
- Plosker SM, Jacobson W, Amato P. Predicting and optimizing success in an intrauterine Insemination programme. Hum Reprod. 1994 Nov; 9 (11): 2014-21.
- Tomlinson MJ, Amissah-Arthur JB, Thompson KA, Kasraie JL, Bentick B. Prognostic indicators for intrauterine insemination (IUI): statistical model for IUI success. Hum Reprod. 1996 Sep; 11 (9): 1892-6.



# Utilizing Data Mining Techniques for Investigating Factors Influencing the Failure of Intrauterine Insemination Infertility Treatment

Abolmasum F<sup>1</sup>/ Alizadeh S<sup>2</sup>/ Asghari M<sup>3</sup>

## Abstract

**Introduction:** Infertility is one of the problems that has caused a lot of psychological and worldly costs on infertile couples. Intrauterine Insemination (IUI) is one of the medically-Assisted Reproduction Techniques (ART) to help infertile couples to have a successful pregnancy. Because of unpredictable results of this technique, identifying the factors influencing the effectiveness of IUI is important. The aim of this study was to identify factors influencing the failure of IUI using data mining techniques.

**Methods:** By utilizing K means algorithm, a descriptive technique of data mining, and Davis-Buldin index, the patients were divided into seven clusters and the features of each cluster were analyzed.

**Results:** Increasing age, overweight, obesity, length and type of infertility in women appeared to be effective factors which were revealed by cluster analysis and investigation of the features of each cluster. Male factors including duration of infertility and spermogram type were other causes of failure in this method of infertility treatment.

**Discussion:** By analyzing the results of clustering technique, the effective factors in the failure of IUI treatment in infertile couples were identified. The obtained results of clustering technique with the consultant of experts can be used for predicting the result of IUI treatment and helping researchers, physicians and infertile couples to choose the best treatment.

**Key words:** Infertility, Intrauterine Insemination, Data Mining, K means

• Received: 20/July/2013 • Modified: 4/Nov/2013 • Accepted: 20/Nov/2013

1. MSc Student of Information Technology, Faculty of Industrial Engineering, K.N Toosi University of Technology, Tehran, Iran; Corresponding Author (faranakabolmasum@yahoo.com)
2. Assistant Professor of Information Technology Department, Faculty of Industrial Engineering, KN Toosi University Of Technology, Tehran, Iran
3. MSc of Information Technology Faculty of Industrial Engineering, K.N Toosi University of Technology, Tehran, Iran