

## *Assessing the Effects of Infertility Treatment Drugs Using Clustering Algorithms and Data Mining Techniques*

Hakimeh Ameri<sup>1</sup>  
Somayeh Alizadeh<sup>2</sup>,  
Marjan Hadizadeh<sup>1</sup>

<sup>1</sup> MSc in information technology, Department of Industrial Engineering, Khaje Nasir Toosi University of Technology, Tehran, Iran

<sup>2</sup> Assistant Professor, Department of Industrial Engineering, Khaje Nasir Toosi University of Technology, Tehran, Iran

(Received June 5, 2013 ; Accepted Jun 5, 2014)

### *Abstract*

**Background and purpose:** The rate of infertility has increased throughout the world. Data mining is a new method for analyzing information from databases. Few studies are done regarding infertility and using data mining in describing and predicting different treatment methods and factors influencing these methods. This paper proposes a model for evaluating the efficacy of different drugs in treatment of infertility among patients treated with IUI.

**Material and Methods:** The records of 26,035 infertile patients (from 1998 to 2009) in Sarim Hospital have been examined. Clinical data of patients were analyzed through data mining methods (Clementine V.12.0). To identify the factors influencing the efficacy of drugs classified data mining and clustering algorithms were used

**Results:** We identified the characteristics of patients with successful treatment using K-Means clustering. CHAID decision tree helped to indicate the result of different drugs in infertility treatments. The proposed model can predict the result of used drugs with 71 percent accuracy.

**Conclusion:** : Data mining techniques can improve the process of treatment in infertile patients by detecting the factors affecting the course of treatment.

**Keywords:** : Infertility, infertility drugs, IUI treatment, data mining

J Mazand Univ Med Sci 2014; 24(114): 26-35 (Persian).

# بررسی میزان تأثیر داروهای درمان ناباروری در بیماران نابارور با استفاده از الگوریتم خوشه بندی و تکنیک های داده کاوی

حکیمه عامری<sup>۱</sup>

سمیه علیزاده<sup>۲</sup>

مرجان هادیزاده<sup>۳</sup>

## چکیده

**سابقه و هدف:** آمار ناباروری در جهان و به خصوص در کشور ما افزایش یافته است. داده کاوی رشته علمی جدیدی در زمینه بازیابی اطلاعات از پایگاه داده ها می باشد. مطالعات معدودی در زمینه ناباروری و استفاده از دانش داده کاوی برای توصیف و پیش بینی روش های مختلف درمانی و شناسایی ویژگی های تأثیر گذار بر موفقیت روش درمان، ارائه شده است. هدف این مقاله ارائه مدلی است برای ارزیابی میزان تأثیر داروهای مختلف درمان ناباروری بر بیمارانی که تحت درمان با روش IUI قرار گرفته اند.

**مواد و روش ها:** برای انجام این تحقیق پرونده های سال ۱۳۷۶ تا ۱۳۸۸ بیماران نابارور بیمارستان صارم بررسی شده اند که ۲۶۰۳۵ رکورد اولیه استخراج شده است. اطلاعات موجود در پرونده بیماران با استفاده از روش های داده کاوی در نرم افزار CLEMENTINE 12.0 تجزیه و تحلیل شدند. برای شناسایی عوامل مؤثر بر تأثیر داروها بر موفقیت درمان از الگوریتم های خوشه بندی و دسته بندی داده کاوی استفاده شد.

**یافته ها:** با استفاده از خوشه بندی K-Means، ویژگی های بیمارانی که درمان آن ها موفق بوده است، شناسایی شده اند. به کمک درخت تصمیم CHAID تأثیر داروهای مختلف درمان ناباروری در بیماران نابارور و در نتیجه باروری مشخص شده است. مدل ساخته شده می تواند نتیجه استفاده از داروی پیشنهادی را با دقت ۷۱ درصد پیش بینی کند.

**استنتاج:** با استفاده از تکنیک های داده کاوی می توان روند درمان بیماران نابارور را بعد از بررسی ها و انجام معاینات و آزمایشات گوناگون و انتخاب فیله های تأثیر گذار بر بیماری، بهبود بخشید.

**واژه های کلیدی:** ناباروری، داروهای ناباروری، روش درمانی IUI، داده کاوی

## مقدمه

روان شناختی، کارکردهای خانوادگی، شغلی و ارتباطی را متأثر می سازد. نتایج پژوهش ها نشان داده اند که ناباروری با افسردگی، استرس، اضطراب، نارضایتی از زندگی و دیگر مشکلات روان شناختی همراه است.

ناباروری به عدم وقوع بارداری با وجود ۱۲ ماه مقاربت صحیح بدون استفاده از روش های جلوگیری اطلاق می شود و نازایی نشان دهنده عدم توانایی زوج در باروری است. ناباروری جنبه های مختلف شخصیتی و

**مؤلف مسئول:** سیمیه علیزاده - تهران - دانشکده مهندسی صنایع، گروه مهندسی فناوری اطلاعات دانشگاه خواجه نصیرالدین طوسی

E-mail: s\_alizadeh@kntu.ac.ir

۱. کارشناس ارشد فناوری اطلاعات، تجارت الکترونیک، دانشکده مهندسی صنایع، گروه مهندسی فناوری اطلاعات دانشگاه خواجه نصیرالدین طوسی تهران

۲. استادیار دانشگاه خواجه نصیرالدین طوسی، دانشکده مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی تهران

۳. کارشناس ارشد فناوری اطلاعات، تجارت الکترونیک، دانشکده مهندسی صنایع، دانشگاه خواجه نصیرالدین طوسی تهران

تاریخ دریافت: ۱۳۹۲/۱۰/۱۵ تاریخ ارجاع جهت اصلاحات: ۱۳۹۳/۲/۱۰ تاریخ تصویب: ۱۳۹۳/۳/۱۵

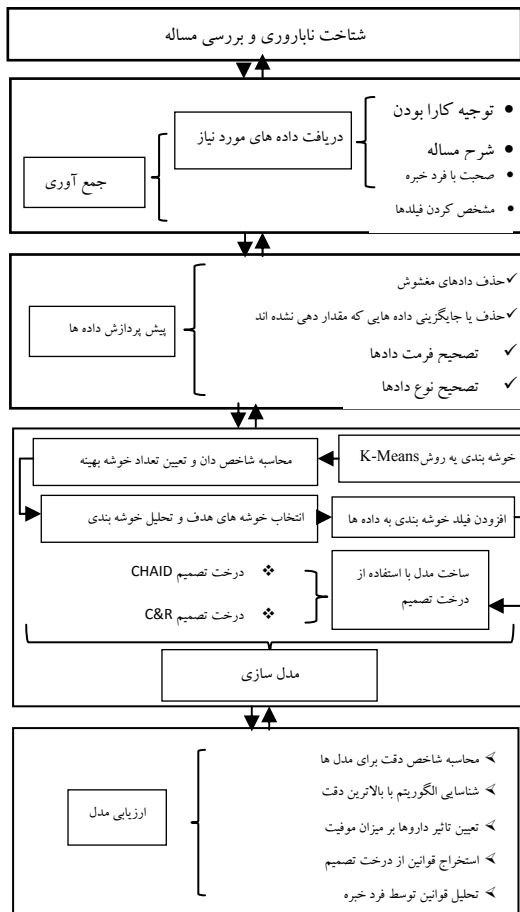
آمار ناباروری در جهان و به خصوص در کشور ما افزایش یافته است. طبق آمارگیری‌هایی که بین ۱۷ هزار زوج ایرانی در سراسر کشور انجام گرفت، مشخص شد که حدود ۲۰ درصد زوجین ایرانی نابارورند که این میزان، از میانگین جهانی بالاتر است. اما براساس گزارش سازمان بهداشت جهانی، شیوع ناباروری در کل دنیا ۱۲ تا ۱۵ درصد است (۱). با توجه به پیشرفت‌های مختلف در عرصه‌های گوناگون دانش پزشکی، امروزه پدیده ناباروری دیگر به عنوان یک بیماری غیر قابل درمان مطرح نیست. پزشکان با کمک تکنیک‌های گوناگون در کنار دانش پزشکی توانسته‌اند با استفاده از داروها و روش‌های کمک باروری در بسیاری از موارد بیماران را درمان کنند و بارداری حاصل گردد. درمان ناباروری در دو دهه‌ی اخیر رشد شگرفی داشته است. درمان ناباروری براساس علت آن انجام می‌گیرد. ناباروری و شناخت درست و کامل پاتوفیزیولوژی ناباروری اهمیت فراوانی در انتخاب روش‌های درمانی توسط پزشک دارد (۲).

قاضی مقدم و همکاران در سال ۱۳۸۸ به بررسی داروی کلومیفن و عوارض درمان این دارو و مقایسه با داروی لتروزول پرداختند و نتیجه گرفتند که لتروزول می‌تواند جایگزین مناسبی برای کلومیفن باشد؛ و عوارض درمان با کلومیفن را نیز ندارد (۸). زینال زاده و همکاران در سال ۱۳۸۷ به مقایسه لتروزول و کلومیفن سترات در تحریک تخمک گذاری در بیماران نابارور مبتلا به سندرم تخمدان پلی کیستیک پرداختند و بیان کردند که لتروزول می‌تواند در حد کلومیفن سترات در تحریک تخمک گذاری و حاملگی در بیماران PCOs مؤثر باشد (۹). معدنی و همکاران (۱۳۸۴) به بررسی عوامل تأثیر گذار بر میزان تخمک گذاری با تاموکسیفن در مبتلایان به عدم تخمک گذاری مزمن، که دارای دوره‌های قاعدگی نامنظم و FSH طبیعی هستند، پرداختند. پیش بینی وقوع تخمک گذاری در نحوه درمان این بیماران مفید بوده و توانایی پزشکان را در

انتخاب داروی مناسب و بی خطر افزایش می‌دهد. نتایج حاصل از این مطالعه نشان می‌دهد که علائم بالینی شامل BMI و هیرسوتیسم، وضعیت آندوکراین و شاخص‌های سونوگرافیک تخمدان زنان اولیگومنوره در مرحله قبل از درمان می‌تواند در پیش‌گویی پاسخ تخمدان‌ها پس از تحریک با داروی تاموکسیفن مؤثر باشد (۱۰). یکی از دانش‌هایی که در کنار علم پزشکی به درمان بیماران کمک زیادی کرده است، داده کاوی و تکنیک‌هایی که در این علم گنجانده شده است می‌باشد. با استفاده از دانش داده کاوی می‌توان به ارزیابی بیماران نابارور پرداخت و به جامعه پزشکان و بیماران در این حوزه کمک کرد تا بتوانند در کوتاه‌ترین زمان و با کم‌ترین هزینه از صحیح‌ترین روش درمانی برای درمان خود استفاده کنند.

پیشرفت‌های بشر در چند دهه اخیر در زمینه جمع‌آوری و ذخیره داده‌ها باعث شده است که ابعاد پایگاه داده‌ها به طور چشمگیری افزایش پیدا کند. صنعت پزشکی جزء معدود صنایعی است که با پایگاه‌های داده‌ای بسیار بزرگی مواجه است. نکته مهم در این پایگاه داده‌های بزرگ، اطلاعات و دانشی است که از آن‌ها استخراج می‌شود و در این زمینه استفاده از روش‌های هوشمند و ساختاریافته برای ارزیابی بسیار ضروری است. داده کاوی رشته علمی جدیدی در زمینه بازیابی اطلاعات از پایگاه داده‌ها می‌باشد. تکنیک‌های مختلفی به لحاظ خطی و غیر خطی برای کاهش ابعاد و رسیدن به ویژگی‌های مفید از یک پایگاه داده وجود دارد که همگی در داده کاوی مطرح است. مطالعات معدودی در زمینه ناباروری و استفاده از دانش داده کاوی برای توصیف و پیش‌بینی روش‌های مختلف درمانی و شناسایی ویژگی‌های تأثیر گذار بر موفقیت روش درمان، ارائه شده است. اما در زمینه بررسی و ارزیابی داروهای درمان ناباروری و استفاده از تکنیک‌های مختلف داده کاوی و بررسی میزان تأثیر و یا پیش‌بینی استفاده از داروهای درمان ناباروری مقاله‌ای

Data Mining است (۱۱) که در تصویر شماره ۱ نمایش داده شده است.



تصویر شماره ۱: مدل پیشنهادی

پس از شناخت مساله ناباروری و روش های درمانی با استفاده از منابع و کتاب های موجود در این زمینه و بررسی ضرورت انجام تحقیق با توجه به رشد ناباروری در کشور و تنوع داروها و روش های درمانی، به کمک افراد خبره و ویژگی ها و فیلدهای مورد نیاز برای انجام این پژوهش انتخاب و جمع آوری شدند. پرونده های بیماران مراجعه کرده به بیمارستان صارم از سال ۱۳۷۶ تا ۱۳۸۸ استخراج شدند. در مرحله پیش پردازش داده ها، روش درمانی (Intrauterine insemination) IUI در این پروژه انتخاب شد که از تعداد ۲۶۰۳۵ داده گرفته شده تحت درمان به روش IUI پس از حذف رکوردهای ناقص و داده های مغشوش، تعداد ۲۵۰۳۵

یافت نشده است. در جدول شماره ۱ تعدادی مقاله در زمینه ناباروری و استفاده از تکنیک های داده کاوی برای بهتر انجام شدن روش های درمانی به طور خلاصه مورد بررسی قرار داده ایم (۷-۳).

جدول شماره ۱: مروری بر مطالعات انجام شده در حوزه ناباروری به کمک داده کاوی

| تکنیک داده کاوی                          | مطالعه کنندگان                                    | نتایج به دست آمده  |
|--|---|--|
| ترکیب الگوریتم ژنتیک و درخت تصمیم        | Güth و همکاران (۲۰۱۰)                             | دقت مدل ترکیبی ایجاد شده ۷۳٫۲ درصد، حساسیت ۷۱٫۶٪، ویژگی ۷۳٫۸ درصد گزارش شده است. مهم ترین پارامتر تاثیر گذار بر موفقیت این روش باروری، سن معرفی شده است.   |
| الگوریتم ژنتیک و برنامه نویسی ژنتیک      | Chen و همکاران (۲۰۰۹)                             | یک روش جدید برای کشف دانش در پایگاه داده (KDD) با رویکرد PSO، و با رمزگذاری یک قاعده تصمیم گیری از موفقیت های جزئی طراحی شده است که می تواند قاعده مطلوب را به شکل یک مدل پیش بینی به دست آورد.  |
| انتخاب ویژگی و شبکه عصبی مصنوعی شبکه یزی | Kumar و Durairaj (۲۰۱۳) و Corani و همکاران (۲۰۱۳) | عواملی با بیشترین ضریب تاثیر شناسایی شدند و مدلی با دقت ۸۴٪ ساخته شده است. یک مدل شبکه یزی برای پیش بینی خروجی عمل IVF معرفی کرده اند که از مدل های شناخته شده الگوریتم های دسته بندی و مستی تخمین بهتری را برآورد میکند.                  |
| طبقه بندی یزی                            | Morales و همکاران (۲۰۰۸)                          | بهترین دقت مدل با استفاده از الگوریتم یزی ساده به میزان ۷۱٫۴۴ درصد بدست آمده است. نتایج به دست آمده به طور عمده متمرکز بر شناسایی در ویژگی اصلی به منظور تفاوت چنین باقیه برای کاشت در رحم هستند، یعنی اندازه بلاستومر و درجه تکه تکه شدن. |

از این رو در این مقاله سعی شده است با استفاده از تکنیک های داده کاوی مدلی برای بررسی میزان تأثیر داروهای درمانی بیماران نابارور ارائه شود. با استفاده از مدل ارائه شده می توان فرآیند درمان بیماران نابارور را سرعت بخشیده و از سوی دیگر اگر عوامل و فاکتورهایی که در مدل بررسی می شوند از سوی افراد خبره مورد تأیید باشد و بر سر انتخاب این عوامل اتفاق نظر بین افراد خبره وجود داشته باشند؛ میزان درستی و دقت مدل در انتخاب دارو برای بیماران بیش تر می شود. ارائه یک چنین مدل های هوشمندی در سیستم های پزشکی به پزشکان و بیماران در زمینه درمان کمک بزرگی خواهد کرد.

## مواد و روش ها

مدلی که برای ارزیابی در این مقاله ارائه شده بر اساس متدولوژی (Cross Industry Process For CRISP)

است. بنابراین تعداد خوشه‌ای که مقدار این شاخص را زیاده‌تر نماید، مقدار بهینه تعداد خوشه‌ها است (۱۳). تعداد خوشه بهینه به دست آمده به کمک شاخص دان، ۹ خوشه است. در ارزیابی و تحلیل خوشه بندی، بررسی می‌شود که در هر خوشه چه رکوردهایی، با چه مقادیری قرار گرفته است، از بین خوشه‌ها کدام خوشه به عنوان خوشه هدف شناخته می‌شود و ویژگی‌های رکوردهایی که در خوشه هدف قرار گرفته‌اند، چیست؟ ویژگی‌های مشترک خوشه‌ها شناسایی می‌شوند. سپس

جدول شماره ۲: پارامترهای انتخاب شده برای تحلیل روش IUI

| ردیف | پارامترهای انتخاب شده   |
|------|---|
| ۱-   | ADTCod (شماره پرونده بیمار)                                       |
| ۲-   | RelationCode (تعداد دفعات مراجعه)                                 |
| ۳-   | سال تولد (زن)   |
| ۴-   | وزن   |
| ۵-   | قد  |
| ۶-   | BMI (kg/m <sup>2</sup> )  |
| ۷-   | Induction of Ovulation (داروهای استفاده شده در طول روش درمان IUI) |
| ۸-   | مدت ناباروری  |
| ۹-   | عامل مردانه (TestMF)  |
| ۱۰-  | عامل تخمک گذاری (TestOF)  |
| ۱۱-  | FemaleFSH (هورمون محرک فولیکول)                                   |
| ۱۲-  | TestTF (عامل مربوط به لوله های فلوپ)                              |
| ۱۳-  | FemaleLH  |
| ۱۴-  | FemaleTSH   |
| ۱۵-  | FemaleTestTesto   |
| ۱۶-  | MaleFSH   |
| ۱۷-  | TestCF  |
| ۱۸-  | BMI Status  |
| ۱۹-  | تعداد اسپرم‌های نرمال بعد از شست و شو                             |
| ۲۰-  | وضعیت رحم (Test CF)   |
| ۲۱-  | سابقه انجام عمل‌های مربوط به باروری                               |
| ۲۲-  | تعداد انجام عمل IUI   |
| ۲۳-  | سابقه بیماری‌های خاص  |
| ۲۴-  | سابقه صرف دارو  |
| ۲۵-  | داروهای مصرف شده  |
| ۲۶-  | سابقه بیماری واگیردار   |
| ۲۷-  | سابقه بیماری‌های روحی   |
| ۲۸-  | ناباروری ثانویه   |
| ۲۹-  | وضعیت نرمال یا غیر نرمال بودن اسپرم                               |
| ۳۰-  | ضخامت آندومتر   |
| ۳۱-  | تعداد اسپرم‌ها بعد از شست و شو                                    |
| ۳۲-  | MaleTSH   |

رکورد به عنوان ورودی وارد نرم‌افزار Celementine 12.0 شد. از بین بیمارانی که تحت درمان به این روش قرار گرفته بودند، تعداد ۲۷۰۳ مورد منجر به بارداری شده بود. با توجه به این که پرونده بیمارانی IUI مورد بررسی قرار گرفته بود، تعداد فولیکول‌ها و اسپرم‌ها در یک بازه یکسان قرار داشته و تفاوت چندانی با هم نداشتند و توسط نرم افزار بی‌اهمیت در نظر گرفته شد. فیلهای انتخابی توسط نرم افزار سپس مورد تأیید افراد خبره (تیم پزشکی و پزشکان متخصص بیمارستان) قرار گرفتند. فیلهای انتخابی در جدول شماره ۲ ذکر شده است.

معیار ADTCod در خوشه‌بندی به این دلیل شرکت داده شد که طول درمان بیمارانی بررسی شده است. اگر بیمار چند عمل داشته باشد، از طریق این فیله مشخص می‌شود. در مرحله بعد که مدل سازی انجام می‌شود، داده‌ها برای انجام عمل خوشه بندی وارد مدل K-means می‌شوند. خوشه‌بندی نوعی یادگیری بدون نظارت است که روی دسته‌های از پیش تعریف شده و یا ویژگی خاصی به عنوان هدف تکیه ندارد. در واقع خوشه‌بندی شکلی از یادگیری به وسیله مشاهدات است. با خوشه‌بندی، نمونه‌های مشابه با هم در یک گروه قرار می‌گیرند (۱۲). با استفاده از شاخص دان از بین خوشه‌های متفاوت که به عنوان ورودی به مدل داده شد، خوشه‌بندی بهینه به دست آمد. این شاخص برای داشتن خوشه‌های متمرکز با مرزهای مشخص مورد استفاده قرار می‌گیرد. شاخص دان به صورت زیر محاسبه می‌شود:

$$D = \min_i |1 - nc| \min_j |1 + 1 - nc| \left( \frac{d(c_i, c_j)}{\max_k = 1 \dots nc (diam(c_k))} \right) \quad (1)$$

که در آن  $d(c_i, c_j)$  و  $diam(c_i)$  به صورت زیر محاسبه می‌شوند.

$$D(c_i, c_j) = \min_{x \in c_i, y \in c_j} \{d(x, y)\} \quad (2)$$

$$Diam(c_i) = \max_{x, y \in c_i} \{d(x, y)\} \quad (3)$$

هدف این شاخص حداکثر کردن فاصله درون خوشه‌ای در ضمن حداقل کردن فاصله برون خوشه‌ای است. مقادیر این شاخص هر چه بزرگ‌تر باشد بهتر

مدلی برای انجام دسته‌بندی ارائه می‌شود. برای انجام این کار می‌توان از دو گروه داده استفاده کرد:

۱- از داده‌هایی که به عنوان ورودی وارد  $k$ -means شده‌اند.

۲- از نتایج خوشه بندی به عنوان داده‌های ورودی به سیستم.

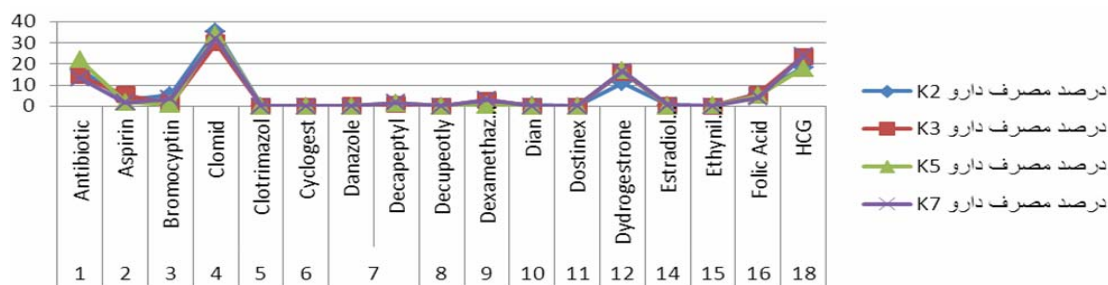
روش دوم یک روش تکنیکال و جدید است که می‌توان به کمک آن دقت دسته بندی را افزایش داد، چرا که پس از انجام خوشه بندی با قرار گرفتن رکوردها در داخل خوشه‌های مناسب، هر رکورد در خوشه‌ای جای گرفته است که از لحاظ ویژگی‌ها به رکوردهای دیگر در همان خوشه نزدیک‌تر است و این می‌تواند به عنوان یک فاکتور مثبت برای ارزیابی به حساب آید. حال بعد از عبور از این مرحله نوبت به ارزیابی درخت دسته بندی با توجه به داده‌های آموزشی که وارد درخت شده است، می‌شود. در مرحله آخر نیز با توجه به نتایج و قوانینی که از مدل استخراج شده است، مدل توسعه داده می‌شود.

برای دسته بندی داده‌ها از درخت تصمیم (Chi-squared automatic interaction detection) CHAID و C&R استفاده شده است. CHAID یک روش دسته‌بندی برای ساخت درخت‌های تصمیم با استفاده از آماره کای دو برای تعیین تجزیه‌های بهینه است. این الگوریتم در ابتدا اهمیت هر یک از متغیرهای پیش بینی کننده را برای پیش بینی متغیر هدف با توجه با آزمون استقلال کای دو تعیین می‌کند. CHAID اغلب به عنوان یک روش اکتشافی استفاده می‌شود و جایگزین مناسبی برای رگرسیون چندگانه خطی و رگرسیون لجستیک است، به ویژه هنگامی که مجموعه ای از داده‌ها برای تجزیه و تحلیل رگرسیون مناسب نیستند. این الگوریتم می‌تواند درختان غیر دودویی را نیز تولید کند. از الگوریتم‌های پرکاربرد دیگر درخت تصمیم، الگوریتم C&R است. این درخت از دسته‌بندی

و رگرسیون برای پیش‌بینی استفاده می‌کند. هر گره ابتدا فیلدهای ورودی را برای یافتن بهترین تجزیه آزمایش می‌کند تا شاخص ناخالصی حاصل از تجزیه کم‌ترین مقدار باشد. تمام تجزیه‌ها دودویی هستند و تا زمانی ادامه پیدا می‌کند که یکی از معیارهای توقف برآورده شود (۱۲، ۱۴). در درخت تصمیم از نتایج خوشه‌بندی استفاده کرده و داده‌های اولیه درخت تصمیم از خروجی روش خوشه بندی است. تغییراتی را روی داده‌ها انجام داده و داده‌ها برای ورود به سیستم آماده می‌شود. برای مثال به بیمارانی که قبلاً از روش‌های کمک باروری استفاده کرده‌اند مقدار یک و به بیمارانی که سابقه انجام روش‌های کمک باروری را ندارند مقدار صفر تعلق می‌گیرد. هدف بررسی میزان تأثیر داروهای مختلف درمان ناباروری در بیماران نابارور است. گره هدف (Target) در درخت تصمیم دارو است که با استفاده از سایر فیلدها مانند مدت ناباروری، سابقه انجام عمل‌های ART، سابقه مصرف داروهای ناباروری، سابقه انجام عمل IUI و نتایج معاینات انجام شده روی بیماران مانند TestCF، TestOF، TestMF و TestTF درخت ارزیابی می‌شود. دقت مدل‌های ساخته شده در دو درخت با هم مقایسه شده و نتایج و قوانین درختی که بالاترین دقت را دارد، استخراج و تحلیل می‌شوند.

## یافته‌ها

با استفاده از الگوریتم خوشه بندی به روش K-Means، خوشه بندی برای ۳ تا ۱۲ خوشه انجام می‌گیرد. پس از محاسبه شاخص دان برای انتخاب خوشه بندی بهینه از میان خوشه‌ها و با توجه به نتایج شاخص دان، خوشه بندی با  $K=9$  به عنوان خوشه بهینه انتخاب می‌شوند. در این خوشه با توجه به نتایج و هدف از خوشه بندی، خوشه با شماره‌های ۲، ۳، ۵، ۷ (خوشه‌هایی که نتایج باروری در آن‌ها مثبت است) به عنوان خوشه هدف انتخاب می‌شوند. نمودار شماره ۱



نمودار شماره ۱: درصد استفاده از داروهای که بیماران در خوشه‌های هدف مصرف کرده‌اند

جدول شماره ۳: میزان پیش بینی درست و نادرست بودن مدل

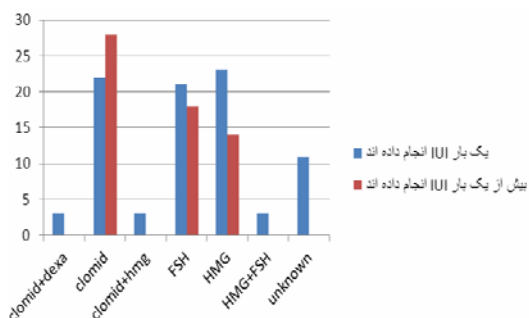
C&R و CHAID

| درخت تصمیم | پیش بینی درست | پیش بینی نادرست |
|------------|---------------|-----------------|
| CHAID      | ۷۰٫۳۸٪        | ۳۰٫۷۱٪          |
| C&R        | ۶۳٫۴۶٪        | ۳۶٫۵۴٪          |

جدول شماره ۴: درصد تأثیر داروهای مختلف در میزان باروری

بیماران

| نام دارو      | تعداد نتایج موفقیت | درصد موفقیت |
|---------------|--------------------|-------------|
| Clomid        | ۷۰                 | ۳۳/۵        |
| Hmg           | ۹۰                 | ۳۹/۲۲       |
| Clomid+Hmg    | ۷۸                 | ۳/۴         |
| Clomid+Dexa   | ۱۲۱                | ۵/۲۷        |
| Natural       | ۸۱                 | ۳/۵۳        |
| Hmg+Fsh       | ۴۸                 | ۲/۰۹        |
| Letrozol      | ۰                  | ۰           |
| Tamioifen     | ۰                  | ۰           |
| Fsh           | ۱۱۲                | ۴/۸۸        |
| Hmg+Fsh+Sf    | ۱۶                 | ۰/۰۷        |
| Hmg+Sf+Bromo  | ۰                  | ۰           |
| Fsh+Sf        | ۰                  | ۰           |
| Deca+Hmg      | ۵                  | ۰/۲۲        |
| Fsh+Bromo     | ۰                  | ۰           |
| Hmg+Fsh+Bromo | ۰                  | ۰           |
| Hmg+Bromo     | ۰                  | ۰           |
| Hmg+Sf        | ۲                  | ۰/۴۴        |
| Unknown       | ۶۲                 | ۲/۷         |



نمودار شماره ۲: درصد موفقیت داروها نسبت به تعداد عمل IUI

انجام شده

شامل لیست داروهای است که بیماران در خوشه‌های هدف از آن‌ها استفاده کرده‌اند و درصد استفاده از هر کدام در هر خوشه مشخص است. همان‌طور که از نمودار نیز مشخص است درصد استفاده از داروها در خوشه‌های موفق تقریباً برابر است.

پس از وارد کردن فیلد نتیجه خوشه بندی به مجموعه داده‌ها و با استفاده از الگوریتم‌های درخت تصمیم CHAID و C&R مدلی برای تأثیر داروهای مختلف بر نتیجه روش درمانی، ساخته می‌شود. با توجه به خروجی‌های به دست آمده در درخت‌های تصمیم، میزان پیش‌بینی‌های درست و نادرست بودن در دو درخت بیان شده است. در اینجا مقایسه بین نتیجه یعنی استفاده از داروی پیشنهادی توسط درخت تصمیم و دارویی که توسط پزشک به بیمار در واقعیت تجویز شده، بیان شده است.

در جدول شماره ۳ پیش‌بینی درست در مدل CHAID نزدیک به ۷۱ درصد موارد دارویی که مدل به بیمار پیشنهاد می‌کند با دارویی که در واقعیت (داده‌های آموزشی) برای بیمار تجویز می‌شود یکسان است؛ و پیش‌بینی نادرست یعنی در ۳۰ درصد موارد خروجی درخت تصمیم با واقعیت یکسان نیست. نتایج حاصل از بررسی درخت CHAID و درصد تأثیر داروهای مختلف در میزان باروری بیماران در جدول شماره ۴ نمایش داده شده است.

نمودار شماره ۲ مقایسه بین میزان موفقیت داروهای مصرفی در بیمارانی که یک بار عمل IUI را انجام داده‌اند و بیمارانی که بیش از یک بار این عمل را انجام داده‌اند را نمایش می‌دهد.

## بحث

آن چه که از نتایج خوشه بندی‌ها با در نظر گرفتن عوامل مختلف مشخص می‌گردد این است که از بین ۹ خوشه ۴ خوشه به عنوان خوشه‌های بهینه‌اند، که در ۲ خوشه از آن‌ها عمل IUI برای بار اول و در ۲ مدل دیگر از آن‌ها عمل IUI برای چندمین بار روی بیماران انجام شده است. بیمارانی که برای اولین بار تحت این عمل قرار گرفته‌اند بیش تر از داروی کلومید استفاده کرده‌اند؛ و HMG در درجه دوم اهمیت است. اما برعکس در خوشه‌هایی که عمل IUI بر روی بیماران برای چندمین بار انجام شده است، از HMG و ترکیبات آن در درجه اول استفاده کرده‌اند و کلومید در درجه دوم مورد استفاده قرار می‌گیرد. در مورد داروهایی که بیماران قبل از انجام IUI از آن‌ها استفاده کرده‌اند نیز جای بحث وجود دارد.

از آن چه که تاکنون بیان شد به این نتایج می‌رسیم که بیمارانی که دارای مشکلات رحمی و لوله‌های فالوپ و مشکل تخمک گذاری هستند و برای اولین بار تحت درمان به روش IUI قرار می‌گیرند، از داروی کلومید و ترکیبات آن برای درمان استفاده می‌کنند. از سوی دیگر بیمارانی که برای بار چندم از IUI برای درمان استفاده می‌کنند، از HMG و ترکیبات آن برای درمان استفاده می‌کنند. حال درصد بیمارانی که برای بار اول از IUI استفاده می‌کنند و بارور می‌شود، ۳۵ درصد از کل بیمارانی می‌باشد که تحت درمان با روش IUI باردار شده‌اند و درصد بیمارانی که برای بار چندم از IUI استفاده می‌کنند و بارور می‌شود ۶۳ درصد از بیمارانی هستند که با IUI باردار شده‌اند. شاید بتوان این طور استنباط کرد که تأثیر داروهایی که برای بار دوم در کنار عمل IUI برای بیمار تجویز می‌شود، تقریباً دو برابر بیشتر از داروهایی است که برای بار اول تجویز می‌شوند، حال چرا از این داروها در همان بار اول استفاده نمی‌شود، جای بحث با متخصصین این دانش را می‌طلبد. همان طور که مشخص است درخت تصمیم

CHAID میزان درصد درستی بیشتری نسبت به درخت C&R دارد. در نتیجه این درخت برای دسته‌بندی در مقایسه با C&R مناسب‌تر دیده می‌شود. بنابراین با توجه به نتایج درخت تصمیم CHAID در مقایسه با درخت C&R، درخت CHAID به عنوان درخت بهینه انتخاب می‌شود. خروجی مدل در اختیار فرد خبره قرار گرفت و پس از بررسی قوانین، نتایجی به صورت زیر استخراج شده است.

✓ تحریک تخمک گذاری به کمک تکنیک (IUI) و استفاده از داروی کلومیفن روش مناسبی برای تحریک تخمک گذاری است.

✓ کلومیفن سترات اولین دارو برای القای تخمک گذاری است اما با این حال بعضی از بیماران به درمان با این دارو به دلایل مختلف پاسخ مثبت نشان نمی‌دهند.

✓ در مواردی که بیماران مبتلا به سندرم تخمدان‌های پلی کیستیک می‌باشند به درمان با کلومیفن پاسخ متفاوت نشان می‌دهند. بررسی‌های پزشکی نشان می‌دهند که استفاده از متفورمین به همراه HMG با دوز پایین، نسبت به استفاده از HMG به تنهایی، می‌تواند موجب رشد فولیکولی و افزایش میزان باروری گردد.

✓ داروی HMG در بیمارانی که تخمدان پلی کیستیک دارند و به درمان با کلومیفن پاسخ نداده‌اند به کار می‌رود، چرا که گاهی اوقات بیماران به کلومیفن پاسخ نمی‌دهند با این که تخمک گذاری هم دارند اما با این وجود بارداری حاصل نمی‌شود. بنابراین می‌توان این طور نتیجه گرفت که ممکن است ناباروری به علت نامشخصی باشد، در این موارد می‌توان برای درمان از HMG استفاده کرد.

✓ استفاده از متفورمین و HMG با دوز پایین ضمن تحریک تخمک گذاری از عوارضی نظیر تحریک بیش از حد تخمدان و چند قلوبی می‌کاهد.

✓ گاهی اوقات HMG در مردانی که مشکل غددی در هیپوفیز و هیپوتالاموس دارند و در نتیجه باعث کاهش کیفیت و کمیت در اسپرم‌ها می‌شود، به کار می‌رود.



لتروزول در درمان ناباروری در مقایسه با کلومیفن، می تواند جانشین مناسبی باشد و هم چنین عوارض منفی ناشی از استفاده کلومیفن را نیز نداشته باشد. اما در این مقاله با توجه به این که داده هایی که در مدل از آن ها استفاده شده فاقد داروی لتروزول می باشند، این عامل در بررسی ها لحاظ نشده است.

مقاله ای در زمینه بررسی تاثیر داروهای ناباروری و انتخاب نوع عمل جراحی به کمک تکنیک های داده کاوی پیدا نشد. شیرین قاضی زاده و همکاران (۸) در مقاله ای با عنوان "مقایسه دو روش تحریک تخمک گذاری [لتروزول + HMG] و [HMG + کلومیفن] در زنان نابارور دارای تخمدان پلی کیستیک و تحت درمان با روش تلقیح داخل رحمی اسپرم" نشان دادند که لتروزول می تواند جایگزین مناسبی برای کلومیفن باشد؛ و عوارض درمان با کلومیفن را نیز ندارد. در پژوهش دیگری که در این زمینه توسط مهتاب زینال زاده و همکاران (۹) با عنوان "مقایسه لتروزول و کلومیفن سیرتات در تحریک تخمک گذاری در بیماران نابارور مبتلا به سندرم تخمدان پلی کیستیک" انجام شده است، اثبات شده است که لتروزول می تواند در حد کلومیفن سیرتات در تحریک تخمک گذاری و حاملگی در بیماران PCOS مؤثر باشد. هم چنین معدنی و همکاران (۱۰) در مقاله ای با عنوان "بررسی عوامل تأثیر گذار بر میزان تخمک گذاری با تاموکسی فن در مبتلایان به عدم تخمک گذاری مزمن، دارای دوره های قاعدگی نامنظم و FSH طبیعی هستند" نشان دادند که علائم بالینی شامل BMI و هیرسوتیسم، وضعیت آندوکراین و شاخص های سونوگرافیک تخمدان زنان اولیگو منوره در مرحله قبل از درمان می تواند در پیش گویی پاسخ تخمدان ها پس از تحریک با داروی تاموکسیفن مؤثر باشد. نتایج حاصل از پژوهش حاضر نشان داد که لتروزول در درمان ناباروری در مقایسه با کلومیفن، می تواند جانشین مناسبی باشد و هم چنین عوارض منفی ناشی از استفاده کاپلومیفن را نیز نداشته

باشد. با توجه به بررسی های انجام شده، نتیجه به دست آمده از این تحقیق کاملاً منطبق با نتیجه کار قاضی زاده و همکاران و معدنی و همکاران میباشد. در هر دو مقاله فوق نتیجه گرفته شده است که لتروزول می تواند جایگزین مناسبی برای کلومیفن باشد. با استفاده از آن چه بیان شد می توان برای داده های جدیدی که وارد سیستم می شوند با در نظر گرفتن فاکتورهای کلیدی که در ساخت درخت از آن ها استفاده شده از مدل استفاده کرد. لازم به ذکر است که در مورد این که از چه داده هایی می توان برای ارزیابی در مدل استفاده کرد تا نتیجه مطلوب به دست آید، نیازمند دانش فرد خبره در این زمینه است. بررسی ها نشان می دهد که می توان از مدل CHAID برای تعیین نوع دارو با توجه به شرایط بیماران استفاده کرد و روند درمان بیماران را تسریع بخشید. با استفاده از مدل ارائه شده اگر بیماران در ابتدا از HMG استفاده می کردند، نتیجه بارداری مثبت حاصل می شد که این امر تسریع در مدت درمان را نشان می دهد. با توجه به مطالعاتی که انجام شد و با توجه به کارهای تحقیقاتی که تاکنون در زمینه داروهای درمان ناباروری انجام شده است می توان به دست آوردهای زیر در آینده اشاره کرد:

- ۱- طراحی سیستمی برای پیش بینی میزان تأثیر استفاده از داروهای باروری در بیماران مبتلا به عدم تخمک گذاری بدون استفاده از سایر روش های کمک باروری
- ۲- طراحی سیستمی برای پیش بینی تأثیر استفاده از داروهای مختلف باروری در میزان موفقیت در باروری.
- ۳- پیش بینی بیان این که کدام یک از داروهای باروری در درمان ناباروری به کمک سایر روش ها (مانند IVF) تأثیر بیشتری دارند.
- ۴- پیش بینی این مورد که استفاده از کدام یک از داروهای درمان ناباروری با توجه به خصوصیات بیماران منجر به سقط جنین می شوند.

## سیاسگزاری

می‌دانند تا از مساعدت و همکاری یکایک این عزیزان تشکر و قدردانی نمایند. هم‌چنین از زحمات و راهنمایی‌های جناب آقای دکتر صارمی کمال تشکر را داریم.

این مطالعه با همکاری پزشکان و پرسنل محترم بیمارستان صارم انجام شده است، محققین بر خود لازم

## References

1. An overview of the latest scientific achievements in the world of Infertility. Available online at : [http:// vimb. basijmed. ir/ Content/Content.aspx?Page Code =39926](http://vimb.basijmed.ir/Content/Content.aspx?PageCode=39926).
2. Guh R, Jackson Wu T, Weng S. Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes. *Expert Systems with Applications* 2011; 38(4): 4437-4449.
3. Saremi A. The definition of infertility, First ed. Tehran: sarem publications 2008(Persian)
4. Chen C, Hsu C, Cheng Y, Li S. Knowledge Discovery on In Vitro Fertilization Clinical Data Using Particle Swarm Optimization. *Bioinformatics and Bioengineering.BIBE'09. Ninth IEEE International Conference 2009 Jun 22-24. Washington, DC, USA; 2009.*
5. Durairaj M, NandhaKumar R. Data Mining Application on IVF Data for the Selection of Influential Parameters on Fertility. *Int J Eng Adv Technol (IJEAT)* 2013; 2(6): 278 – 283.
6. Corani G, Magli C, Giusti A, Gianaroli L, & Gambardella L. M. A Bayesian network model for predicting pregnancy after in vitro fertilization. *Computers in biology and medicine.* 2013, 43(11): 1783-1792.
7. Morales DA, Bengoetxea B, Larrañaga P, García M, Franco Y, Fresnada M, and et.al. Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. *Comput Methods programs Biomed.* 2008; 90(2): 104-116.
8. Ghazizadeh sh, Amooei M, Bagheri M, Ghlichkhani M, Masoomi M. Comparing the Effects of Clomiphene-HMG and Letrozole-HMG on Ovulation Induction in Infertile Women. *J Reprod Infertil* ,2009; 10(2):109-114. (Persian).
9. Zeinalzadeh M, Esmailpour M, Nazari T. Comparison of Letrozole and Clomiphene Citrate for induction of ovulation in patients with polycystic ovarian syndrome. *JBUMS* 2008; 10(3): 35-41. (Persian).
10. Madani T, Zaferani F, Peyrovan S, Eshrati B. Identification of influential factors on ovarian responsiveness with tamoxifen in type-2 WHO anovulatory patients. *J Mazandaran Univ Med Sci* 2005; 15 (48):65-72. (Persian).
11. Han J, Kamber M. chapter 1. Introduction: Data Mining. *Concepts and Techniques.* 2nd ed. San Francisco: Morgan Kaufman Publisher. 2006. P 1-45.
12. Alizadeh S, Ghazanfari M, Teimorpour B *Data Mining and Knowledge Discovery*, 2nd ed. Tehran: Publication of Iran University of Science and Technology. 2011 (Persian).
13. Ilc' N, Modified Dunn's cluster validity index based on graph theory. *Przeglad Elektrotechniczny* 2012; 88(2): 126-131.
14. Alizadeh S, Mlek mahmodi S. *Data mining and knowlede discovery,step by step.* Tehran: KNTU university publication; 2011(Persian).