مجله دانشگاه علوم پزشکی قم
دوره نهم، شماره سوم، خرداد ۹۴
صفحه ۴۹ الی ۵۹

مقاله پژوهشی
(Original Article)

تشخیص ناباروری به کمک تکنیک‌های داده‌کاوی

سلبی حیدری[1*]، ابوالفضل سعیدی‌فر[2]

چکیده

زمینه و هدف: با توجه به جمع‌آوری حجم بالای داده‌ها در مراکز درمانی و استفاده مناسب از آنها جهت کشف و تشخیص بیماری، نیاز به علم و ابزاری که روی این داده‌ها تحلیل انجام دهد، لازم و ضروری است. در این مطالعه تشخیص ناباروری به کمک تکنیک‌های داده کاوی مورد بررسی قرار گرفت.

روش بررسی: در این مطالعه اطلاعات از میان پرونده‌های مراجعه‌کنندگان مرکز درمان ناباروری جهاد دانشگاهی قم استخراج و ۷۰۰ نمونه از میان ۱۴۲۴۲ پرونده موجود در ۱۵ سال انتخاب شد و از متغیرهای سن، مدت ناباروری، نسبت خانوادگی، سابقه ناباروری خانوادگی، شغل مرد، نوع سیکل قاعدگی زن، هیرسوتیسم، گالاکتوره، آمنوره، نوع ناباروری، شاخص توده بدنی زن، مصرف دخانیات و آزمایشهای اسپرم مرد استفاده شد. جهت پیش‌بینی از الگوریتم‌های C5.0، C&R tree، CHAID و خوشه‌بندی از الگوریتم K-means و برای تعیین تعداد خوشه بهینه از شاخص دیویس - بولدین استفاده گردید.

یافته‌ها: با توجه به مدل مورد قبول الگوریتم CHAID که خطای کمتری دارد، مهم‌ترین عوامل تأثیرگذار در ناباروری به ترتیب شاخص توده بدنی زن، سن زن، بیماری هیرسوتیسم، ناباروری خانوادگی، بیماری گالاکتوره، مقدار اسپرم مرد در هر میلی‌لیتر، مدت ناباروری، سن مرد و نسبت فامیلی زوجین بود. طبق این مدل، عوامل زنانه مهم‌تر از عوامل مردانه شناسایی شد.

نتیجه‌گیری: در نتایج این پژوهش، تأثیر بیشتر عوامل زنانه در ناباروری پیش‌بینی شده است.

کلید واژه‌ها: ناباروری؛ داده کاوی؛ الگوریتم K-means؛ درخت تصمیم.

لطفاً به این مقاله به‌صورت زیر استناد نمایید:

Heydari S, Saeedifar A. Infertility diagnosis by data mining techniques. Qom Univ Med Sci J 2015;9(3):49-59. [Full Text in Persian]

[1]دانشجوی کارشناس ارشد آمار اجتماعی - اقتصادی، واحد اراک، دانشگاه آزاد اسلامی، دانشگاه علوم پزشکی قم، قم، ایران.

[2]استادیار ریاضی و آمار، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران.

[*]نویسنده مسئول مکاتبات:
سلبی حیدری، واحد اراک، دانشگاه آزاد اسلامی، دانشگاه علوم پزشکی قم، قم، ایران؛

آدرس پست الکترونیکی:
heydari@muq.ac.ir

تاریخ دریافت: ۹۳/۵/۱۸

تاریخ پذیرش: ۹۳/۸/۲۰

مقدمه

ناباروری به عدم توانایی در باردارشدن پس از یک‌سال نزدیکی، بدون جلوگیری گفته می‌شود که به‌طورکلی ۵۰٪ مربوط به فاکتور زنانه، حدود ۳۵٪ فاکتور مردانه و ۱۵٪ بدون علت است. زوج‌ها باید بدانند که وقوع حاملگی نیاز به زمان داشته و در هر دوره تخمک‌گذاری، ۳۰٪ شانس حاملگی وجود دارد (۱). باتوجه به اینکه داشتن فرزند و برخورداری از لذت پدر و مادر شدن حق هر مرد و زن متأهل است، اما نباید فراموش کرد که حدود ۱۰-۸٪ زن و شوهرها با وجود تمایل به داشتن فرزند، بچه‌دار نمی‌شوند (۲). ناباروری یک بحران پیچیده زندگی است که از لحاظ روانشناختی تهدیدکننده و از جنبه هیجانی نیز ناتوان‌کننده است و مانند سایر فقدان‌ها باعث تحلیل انرژی روانشناختی فرد که معمولاً بایستی صرف بهره‌مندی از زندگی و حل مشکلات گردد، می‌شود (۳). میزان ناباروری در ایران ۲۲-۹٪ تخمین زده شده است (۴). طبق گزارش سازمان بهداشت جهانی، ۸۰ میلیون داوطلب درمان ناباروری در دنیا وجود دارد (۵). باوجود تمام پیشرفت‌های علمی، کماکان درصد قابل‌توجهی از زوج‌ها در ردیف کسانی قرار می‌گیرند که باوجود تلاش و تدابیر مختلف پزشکی همچنان در درمان ناباروری آنان توفیقی حاصل نشده است (۶). در طی ۲ دهه گذشته در بررسی نازایی سه تغییر عمده رخ داده است:

۱. تغییر و تحول در باروری مصنوعی و سایر تکنولوژی‌های باروری کمک‌کننده؛
۲. آگاهی بیشتر جمعیت از درمان‌های موجود برای نازایی؛
۳. افزایش تعداد زنان بالای ۳۵ سال که در جستجوی خدمات درمانی برای نازایی هستند (۷).

در زنان، اختلالات قاعدگی، عدم تخمک‌گذاری و سقط خودبه‌خودی در اثر کار فیزیکی سنگین، کار در اتاق عمل (بیهوشی)، دندانپزشکی و عوارض شغلی دیگر مانند مواجهه با مواد ضدعفونی‌کننده، فلزات سنگین، حلال‌های شیمیایی و ... اتفاق می‌افتد (۸). نمایه توده بدنی بیش از ۳۰، خطر تأخیر باروری را افزایش می‌دهد. از سوی دیگر، لاغری بیش از حد؛ یعنی نمایه توده بدنی کمتر از ۲۰ نیز می‌تواند احتمال باروری را کاهش دهد (۹).

همچنین پذیرش رحم و کیفیت تخمک با افزایش سن کاهش می‌یابد (۱۰). براساس مطالعات انجام‌شده در بیماران پلی‌کیستیک، افزایش سن و چاقی از عوامل خطر عمده در ایجاد هیپرپلازی آندومتر محسوب می‌شود (۱۲،۱۱) که باعث ناباروری می‌گردد. در یک بررسی کاهش واضح در باروری بعد از ۳۵ سالگی گزارش شده است (۱۳).

داده‌کاوی با پرداختن به استخراج اطلاعات مفید و دانش از حجم زیاد داده‌ها، الگوهای حاوی اطلاعات را در داده‌های موجود جست‌وجو می‌کند. این الگوها و الگوریتم‌ها می‌توانند توصیفی باشند؛ بدین معنی‌که داده‌ها را توصیف و یا جنبه پیش‌بینی داشته باشند و از متغیرها برای پیش‌بینی ارزش‌های ناشناخته سایر متغیرها استفاده کنند. داده‌کاوی توصیفی درپی یافتن اگرها در فعالیت‌ها یا اقدامات گذشته است و داده‌کاوی پیش‌بینانه با نگاه به سابقه رفتار، آینده را پیش‌بینی می‌کند (۱۴). حوزه ناباروری در پزشکی از جمله حوزه‌هایی است که می‌توان با به کارگیری داده‌کاوی در آن، به نتایج بسیار مفیدی دست یافت (۱۵). تاکنون مطالعات متعددی در حوزه کاربرد تکنیک‌های داده‌کاوی 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 IVF (In Vitro Fertilization) و یا کاربرد تکنیک‌های هوش مصنوعی برای پیش‌بینی نتایج همین روش درمانی اشاره نمود (۱۶). در پژوهش حاضر از طریق داده‌کاوی، مدل و الگویی ارائه شده است که می‌تواند به پرسش‌های زیر پاسخ دهد:

آیا با استفاده از داده‌کاوی می‌توان ناباروری را تشخیص داد؟

مدل بهینه برای تشخیص ناباروری چیست؟

روش بررسی

مطالعه حاضر به‌صورت داده محور روی داده‌های ناباروری موجود در پرونده‌های مرکز فوق‌تخصصی درمان ناباروری جهاد دانشگاهی قم از سال ۱۳۹۲- ۱۳۷۸ اجرا شد. جامعه آماری تحقیق مورد نظر شامل تمامی زوجینی بود که به مرکز فوق مراجعه و دچار ناباروری بودند. در این مرکز تعداد کل پرونده‌های موجود از سال ۱۳۷۸تا آخر سال ۱۳۹۲ برابر ۱۴۲۴۲ پرونده بود.

با توجه به عدم وجود پرونده الکترونیک، همچنین در دسترس نبودن اطلاعات کامل از بیماران و جمع‌آوری تمام اطلاعات به‌علت مشکلات زیاد در پرونده بیماران مراجعه‌کننده، یک نمونه ۷۰۰ تایی از بین کل پرونده‌هایی که دارای داده‌های مناسب بود انتخاب شد. پس از جمع‌آوری، اطلاعات توسط اساتید آماری و مامایی مورد تأیید قرار گرفت.

در ادامه، نمونه‌ها به کمک نمونه‌گیری طبقه‌بندی (از هر سال به نسبت تعداد پرونده‌های موجود در آن سال) انتخاب شدند و متغیرهای مورد نیاز ثبت گردید که این تعداد به ترتیب در جدول شماره ۱ ارائه شده است.

جدول شماره ۱: تعداد پرونده‌ها و نمونه‌ها

سال	پرونده موجود	نمونه
	تعداد	تعداد
۱۳۷۸	۱۸۰	۱۰
۱۳۷۹	۵۱۰	۲۵
۱۳۸۰	۶۵۳	۳۲
۱۳۸۱	۵۶۳	۲۸
۱۳۸۲	۵۵۵	۲۷
۱۳۸۳	۶۷۴	۳۳
۱۳۸٤	۸۵۴	۴۲
۱۳۸۵	۹۸۵	۴۸
۱۳۸٦	۱۰۶۰	۵۲
۱۳۸۷	۱۰۵۷	۵۲
۱۳۸۸	۱۱۲۱	۵۵
۱۳۸۹	۱۳۰۹	۶۴
۱۳۹۰	۱۴۰۱	۶۹
۱۳۹۱	۱۶۶۲	۸۲
۱۳۹۲	۱۶۵۸	۸۱

فرآیند اجرایی تحقیق براساس مراحل مدل شرح داده شده است: بیماران به‌منظور درمان‌های مختلف به این مرکز مراجعه کرده بودند و اطلاعات شخصی زوجین و نتایج آزمایشها در پرونده‌های آنها ثبت و نگهداری شده بود. از بین فیلدهای موجود در پرونده‌ها، برخی فیلدهای مؤثرتر که ۱۶ عدد بود، براساس مطالعات گذشته در این زمینه و نیز مشورت با متخصصین رشته زنان، زایمان و مامایی به دست آمده، انتخاب و به کار گرفته شد که عبارت بودند از:

سن زن، سن مرد، شغل مرد، نسبت خانوادگی، مدت زمان ناباروری، نوع ناباروری، نظم قاعدگی، آمنوره، گلاکتوره، هیرسوتیسم، سابقه ناباروری خانوادگی، مصرف دخانیات، شاخص توده بدنی زن، تعداد اسپرم در هر میلی‌لیتر، تعداد کل اسپرم و تعداد اسپرم طبیعی.

در این بررسی با توجه به موضوع تحقیق در رابطه با داده‌کاوی از نرم‌افزار Clementine نسخه ۱۲ استفاده شد.

در مرحله اول و دوم، قبل از استفاده از نرم‌افزار جهت پاکسازی و یک‌پارچه‌سازی داده‌ها، کارهایی از قبیل از بین بردن داده‌های دارای اختلال و یک‌پارچه‌سازی فایل‌های متفاوت و یکسان‌سازی داده‌ها انجام شد که در مرحله یک‌پارچه‌سازی و تبدیل داده‌ها، اطلاعاتی که به‌صورت پاسخ تشریحی از مراجعه‌کنندگان جمع‌آوری شده بود، به‌صورت کدگذاری مشخص گردید.

در مرحله سوم (مرحله انتخاب داده‌ها)، داده‌های مرتبط از بین مجموعه اطلاعات انتخاب شد.

در مرحله چهارم (مرحله انتقال داده‌ها)، در بخش Sources از منبع Excel استفاده گردید.

در مرحله پنجم (مرحله داده‌کاوی) با استفاده از تکنیک‌ها و الگوهای مناسب و مفید، مدل‌های C&R Tree، C5.0 و CHAID برای پیش‌بینی و مدل K-means جهت خوشه‌بندی داده‌ها به کار برده شد.

الگوریتم C5.0: این الگوریتم برای ساخت یک درخت تصمیم یا یک مجموعه قاعده استفاده می‌شود. یک مدل C5.0 به‌وسیله تجزیه داده براساس ویژگی‌ها، حداکثر بهره از اطلاعات را می‌برد. هر زیربخش به بخش‌های کوچکتری تقسیم و مجدداً زیربخش‌ها به بخش‌های کوچکتر تقسیم می‌شوند.

الگوریتم C&R tree: الگوریتم درخت رگرسیون، دسته‌بندی یک روش پیش‌بینی و دسته‌بندی مبتنی بر درخت است که مشابه الگوریتم C5.0 در آن، از روش بخش‌بندی بازگشتی برای تقسیم و آموزش رکوردها در بخش‌ها با مقادیر ویژگی خروجی مشابه استفاده می‌شود. بخش به دو زیربخش تقسیم می‌شود و هریک از زیربخش‌ها به دو زیربخش دیگر و این روند ادامه می‌یابد تا شرط و معیار متوقف گردد. تمام بخش‌ها دودویی هستند.

الگوریتم CHAID: CHAID، مخفف مجذور مربع شناسایی‌کننده تعاملات خودکار است. این الگوریتم مقادیری را که به‌صورت آماری همگن در نظر گرفته شده‌اند با توجه به متغیر

هدف و حفظ تمام مقادیری که ناهمگن هستند، ادغام می‌کند. CHAID یک درخت باینری نیست و می‌تواند بیش از دو دسته در هر سطح خاص از درخت ایجاد کند (۱۷).

در مرحله ششم (مرحله ارزیابی الگو)، الگوی مناسب‌تر انتخاب می‌شود.

مرحله هفتم داده‌کاوی ارائه دانش بوده که در آن دانش کشف‌شده به‌صورت بصری ارائه می‌گردد.

در الگوریتم K-means باید تعداد خوشه‌های بهینه مشخص گردد. برای اینکار شاخص‌های مختلفی وجود دارد. از جمله شاخص دیویس- بولدین که در این شاخص از کمترین عدد به دست آمده، به‌عنوان تعداد K بهینه استفاده می‌شود.

یافته‌ها

آماره‌های توصیفی تحقیق مطابق جدول شماره ۲ می‌باشد.

جدول شماره۲: آمارهای توصیفی

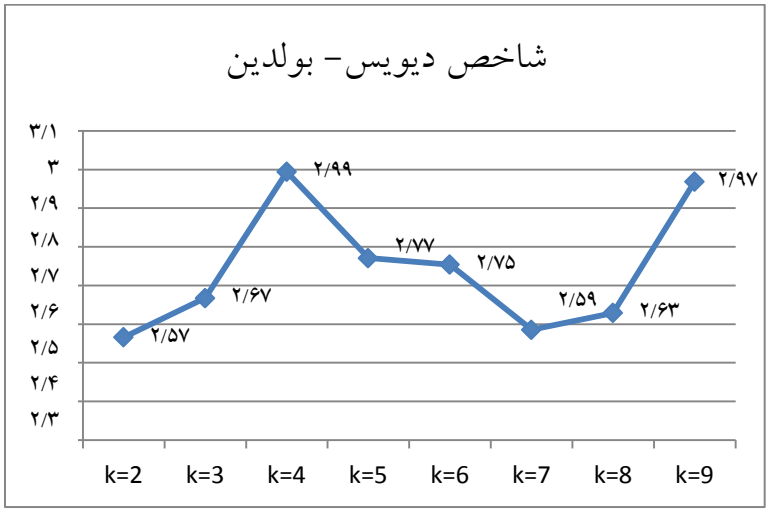
عنوان	میانگین	انحراف معیار	حداقل	حداکثر
سن زن	۲۸	۵/۸	۱۶	۴۶
سن مرد	۳۳	۶/۸	۱۹	۷۲
مدت زمان ناباروری (ماه)	۶۵	۵۵/۸	۲	۳۲۴
مقدار اسپرم در هر میلی‌لیتر	۵۸/۱	۴۸/۹	۰/۰۲	۳۰۰
مقدار اسپرم کل	۱۷۸/۶	۱۵۹/۴	۰/۰۲	۸۴۰
مقدار اسپرم طبیعی	۱۲/۱	۱۱/۸	۱	۹۷

در این مطالعه، میانگین سن زنان ۲۸ سال، کمترین سن آنها ۱۶ سال و بیشترین سن ۴۶ سال بود و میانگین سن مردان ۳۳ سال، بیشترین سن ۷۲ سال و کمترین ۱۹ سال بود. در رابطه با مدت ناباروری، کمترین مدت ۲ ماه و بیشترین آن ۳۲۴ ماه برآورد شد و میانگین مدت ناباروری حدود ۶۵ ماه بود. بیشترین مقدار تعداد اسپرم در هر میلی‌لیتر، ۳۰۰×۱۰^{۶}، کمترین $۰/۰۲\times۱۰^{۶}$ و میانگین آن ۵۸×۱۰^{۶} بود. در تعداد کل اسپرم، بیشترین مقدار ۸۴۰×۱۰^{۶}، کمترین $۰/۰۲\times۱۰^{۶}$ و میانگین آن ۱۷۸×۱۰^{۶} بود.

در مورد تعداد اسپرم طبیعی، میانگین ۱۲٪، کمترین مقدار ۱٪ و بیشترین آن، ۹۷٪ برآورد شد و بیشترین فراوانی به مقدار ۱٪ تعلق داشت (جدول شماره ۲).

برای داده‌های این تحقیق از شاخص دیویس- بولدین مطابق نمودار شماره ۱ استفاده شد.

با توجه به این نمودار مشخص گردید در حالت K=۲، این شاخص کمترین عدد را دارا می‌باشد.



نمودار شماره ۱: نمودار دیویس- بولدین

بیماران در هریک از گروه‌ها دارای ویژگی‌های مشابهی بودند. در جدول شماره ۳ ویژگی‌های هریک از گروه‌ها ارائه شده است.

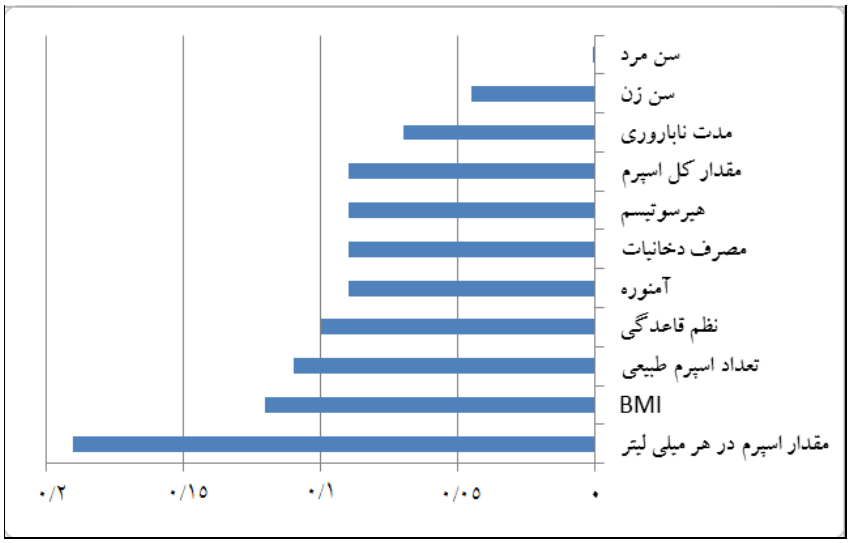
جدول شماره ۳: ویژگی‌های گروه‌های مربوط به مراجعه‌کنندگان مرکز درمان ناباروری

ویژگی	گروه	گروه اول (۴۰۳ مورد) میانگین±انحراف معیار	گروه دوم (۲۹۷ مورد) میانگین±انحراف معیار
میانگین سن زن		۲۹/۱±۶/۱	۲۶/۸±۵/۲
میانگین سن مرد		۳۳/۸±۷/۲	۳۲±۶/۱
		تعداد (درصد)	**تعداد (درصد)**
شغل مرد	مشاغل نشسته	۱۴۱ (۳۴/۸)	۱۰۰ (۳۳/۸)
	مشاغل ساختمانی	۱۰ (۲/۵)	۰
	مشاغل حرارتی	۴ (۱)	۱۰ (۳/۴)
	کار با آهن	۲ (۰/۵)	۷ (۲/۴)
	کار با چوب	۹ (۲/۲)	۵ (۱/۷)
	کار با رنگ	۱۸ (۴/۵)	۷ (۲/۴)
	کشاورز	۸ (۲)	۳ (۱/۰۱)
	مدرس	۷۹ (۱۹/۷)	۶۲ (۲۰/۹۵)
	آزاد و سایر	۱۳۲ (۳۲/۸)	۱۰۳ (۳۴/۵)
		درصد	**درصد**
نسبت خانوادگی	غریبه	٪۶۹/۲	٪۶۹/۷
	فامیل دور	٪۱۱/۲	٪۱۳/۸
	فامیل نزدیک	٪۱۹/۷	٪۱۶/۵
نوع ناباروری	اولیه	٪۸۴/۶	٪۸۲/۲
	ثانویه	٪۱۵/۴	٪۱۷/۸
مدت ناباروری (به ماه)		۶۷/۵	۶۰/۹
آمنوره	مثبت	٪۵/۴۶	٪۴۴/۱۱
	منفی	٪۹۴/۵	٪۵۵/۸۹
گالاکتوره	مثبت	٪۸/۷	٪۱۰/۸
	منفی	٪۹۱/۳	٪۸۹/۲
هیرسوتیسم	مثبت	٪۱۳/۶	٪۳۱/۳
	منفی	٪۸۶/۴	٪۶۸/۷
نظم قاعدگی	منظم	٪۱۰۰	٪۰/۳
	نامنظم	۰	٪۹۹/۷
BMI	۱۸/۵-۱۵ لاغر	٪۳/۸	٪۳/۲
	۲۴/۹-۱۸/۵ وزن طبیعی	٪۴۶/۷	٪۳۶/۴
	۲۹/۹-۲۵ اضافه وزن	٪۳۱/۱	٪۳۲/۰۲
	۳۹/۹-۳۰ چاقی	٪۱۸/۳	٪۲۷/۷
	=>۴۰ چاقی شدید	۰	٪۰/۸
سابقه ناباروری در خانواده	بدون نسبت	٪۷۵/۷	٪۷۲/۴
	فامیل دور مرد	٪۳/۵	٪۳/۰۳
	فامیل دور زن	٪۸/۲	٪۱۰/۸
	فامیل دور هر دو	٪۱/۵	٪۰/۷
	فامیل نزدیک مرد	٪۴/۵	٪۴/۷
	فامیل نزدیک زن	٪۶/۲	٪۷/۴
	فامیل نزدیک هر دو	٪۰/۵	٪۱/۰۱
مصرف دخانیات	اعتیاد به مواد (در مردان)	٪۳/۲	٪۲/۰۲
	ترک مواد (در مردان)	٪۰/۵	٪۱/۰۱
	کشیدن سیگار (در مردان)	٪۱۷/۱	٪۱۲/۱
	کشیدن سیگار و قلیان (در مردان)	٪۰/۵	۰
	کشیدن قلیان (در مردان)	٪۰/۷	٪۱/۰۱
	کشیدن قلیان گاهی (در زنان)	۰	٪۰/۳
	عدم مصرف	٪۷۷/۹	٪۸۲/۸
	کشیدن قلیان (هم در زنان و هم در مردان)	۰	٪۰/۷
تعداد اسپرم در هر میلی‌لیتر (میلیون)		۸۶/۴	۸۲/۶
تعداد اسپرم در کل (میلیون)		۲۵۲/۴	۲۴۳/۵
اسپرم نرمال (٪)		۲۲/۲	۲۲/۷

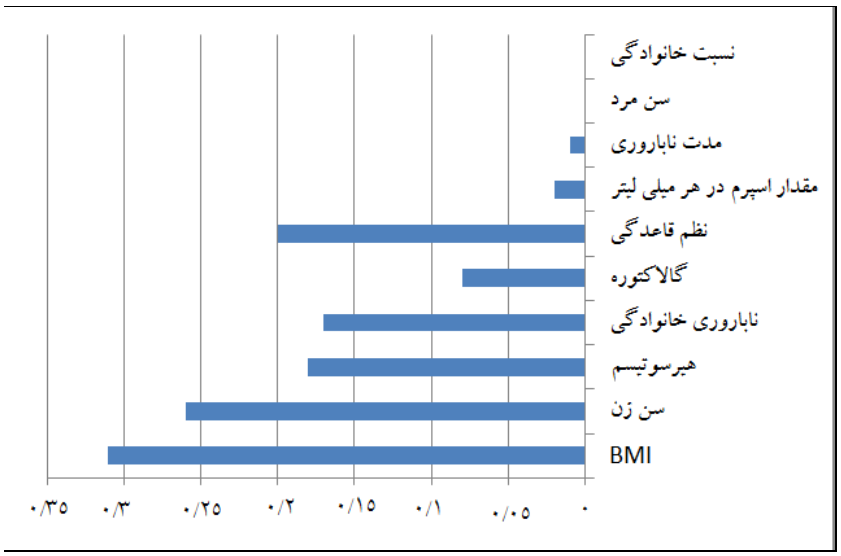
گروه اول بیماران: این گروه بیشترین فراوانی از نظر تعداد را در بین گروه‌ها داشتند، میانگین سن زنان، بالای ۲۹ سال و برای مردان حدود ۳۴ سال بود که از نظر پزشکی برای باروری، سن بالا محسوب می‌شود. از نظر نسبت خانوادگی؛ بیشترین زوجین غریبه بودند. حدود ۸۵٪ زوجین در این گروه دچار ناباروری اولیه بودند و از نظر سابقه ناباروری در خانواده، ۷۵٪ زوجین سابقه‌ای نداشتند و بیشترین درصد زنان در این گروه، دارای وزن طبیعی بودند.

گروه دوم بیماران: در این گروه تقریباً همه زنان دچار نامنظمی قاعدگی (۹۹/۷٪) بودند که ممکن است علت آن داشتن آمنوره و هیرسوتیسم باشد، همچنین حدود ۶۰٪ زنان دچار اضافه وزن و چاقی بودند.

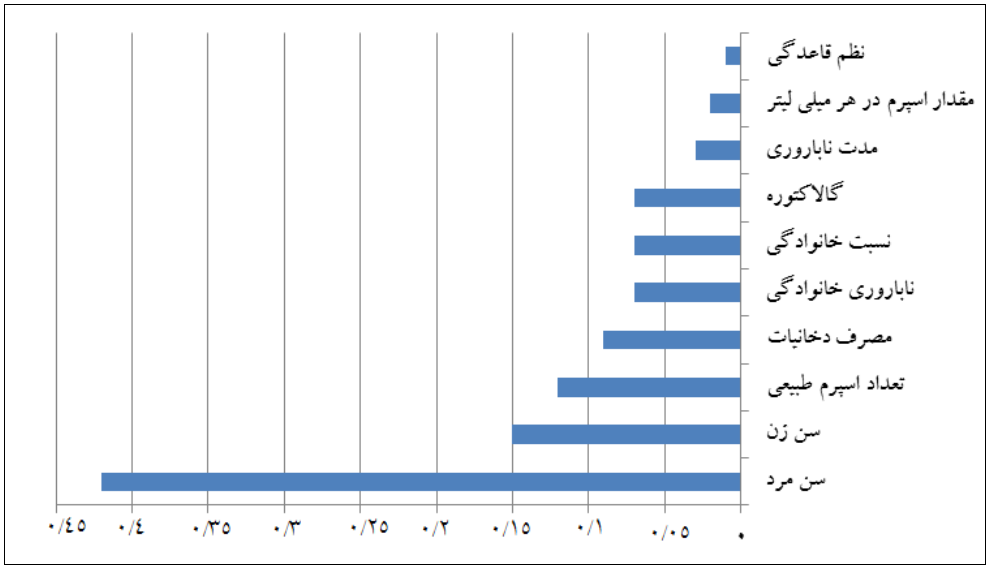
جهت پیش‌بینی ناباروری از الگوریتم‌های درخت تصمیم استفاده گردید. برای شناسایی بهترین مدل، جدول آنالیز آنها بررسی شد تا مدلی که درصد خطای کمتری دارد انتخاب گردد (نمودار شماره ۴-۲).



نمودار شماره ۲: عوامل مؤثر در مدل C&R tree



نمودار شماره ۳: عوامل مؤثر مدل CHAID



نمودار شماره ٤: عوامل مؤثر مدل C5.0

با توجه به جدول شماره ۴، بهترین مدل، مدل CHAID می‌باشد.

جدول شماره ٤: خطای آزمون مدل‌های درخت تصمیم

مدل	C&R tree	CHAID	C5.0
خطا	۱۱/۸۵	۱۰/۷۱	۱۱/۲۸

در مدل C&R tree، بیشترین عامل مؤثر در ناباروری اولیه و ثانویه مقدار اسپرم در هر میلی‌لیتر است. نمایه توده بدنی زن، مقدار اسپرم نرمال، نظم قاعدگی، آمنوره، مصرف دخانیات، هیرسوتیسم و مقدار کل اسپرم، در رتبه‌های بعدی اهمیت قرار دارند. سن مرد نیز دارای کمترین درجه اهمیت است (جدول شماره ۵).

جدول شماره ٥: عوامل مؤثر در الگوریتم‌های درخت تصمیم

مدل / رتبه اهمیت	CHAID	C&R tree	C5.0
۱	BMI	مقدار اسپرم در هر میلی‌لیتر	سن مرد
۲	سن زن	BMI	سن زن
۳	هیرسوتیسم	تعداد اسپرم طبیعی	تعداد اسپرم طبیعی
٤	ناباروری خانوادگی	نظم قاعدگی	مصرف دخانیات
٥	گالاکتوره	آمنوره	ناباروری خانوادگی
٦	نظم قاعدگی	مصرف دخانیات	نسبت خانوادگی
۷	مقدار اسپرم در هر میلی‌لیتر	هیرسوتیسم	گالاکتوره
۸	مدت ناباروری	مقدار کل اسپرم	مدت ناباروری
۹	سن مرد	مدت ناباروری	مقدار اسپرم در هر میلی‌لیتر
۱۰	نسبت خانوادگی	سن زن	نظم قاعدگی
۱۱	–	سن مرد	–

در مدل C5.0، بیشترین تأثیر را سن مرد در ناباروری اولیه و ثانویه دارا می‌باشد. سن زن، مقدار اسپرم طبیعی و مصرف دخانیات در رتبه‌های بعدی قرار دارند، همچنین نظم قاعدگی کمترین تأثیر را داشته است.

در مدل CHAID، بیشترین عامل مؤثر در ناباروری اولیه و ثانویه نمایه توده بدنی زنان است. سن زن، همچنین هیرسوتیسم، ناباروری خانوادگی و گالاکتوره در درجه‌های بعدی اهمیت قرار دارند. نسبت خانوادگی در کمترین درجه اهمیت می‌باشد.

در این مدل، سن مردان به سه گروه زیر ۳۱ سال، ۳۵-۳۱ سال و بالای ۳۵ سال تقسیم شده است که بیشترین تعداد را گروه زیر ۳۱ سال به خود اختصاص می‌دهد. در همه گروه‌ها، بیشترین افراد دچار ناباروری اولیه هستند. در این مطالعه از آنجایی که الگوریتم CHAID به‌عنوان بهترین مدل انتخاب گردید، لذا به بررسی عوامل مؤثر آن پرداخته شد.

در گروه اول (مردان زیر ۳۱ سال)، هیرسوتیسم زن به‌عنوان عامل مؤثر در ناباروری پیش‌بینی شد و در صورتی‌که زن دچار هیرسوتیسم بود، ۱۰۰٪ احتمال ناباروری وجود داشت و اگر زن هیرسوتیسم نبود و دچار گالاکتوره بود به احتمال ۰/۷۵٪ دچار ناباروری می‌شد.

در گروه دوم (گروه سنی ۳۵-۳۱ سال)، پیش‌بینی گردید نمایه توده بدنی زن مؤثر بوده است و به احتمال ۶۴/۵٪ زنان چاق و چاقی شدید در این شرایط سنی دچار ناباروری اولیه و حدود ۳۵/۵٪ دچار ناباروری ثانویه شده بودند.

در گروه سوم (مردان بالای ۳۵ سال) که به این مرکز مراجعه کرده بودند، مدت زمان ناباروری به‌عنوان عامل مهم شناخته شد. در دسته کمتر از ۱۴۴ ماه ناباروری، نظم قاعدگی زن مؤثر بوده است و پیش‌بینی گردید زنان دارای قاعدگی منظم و دارای سن کمتر از ۲۸ سال، به احتمال ۸۱/۸٪ دچار ناباروری اولیه بوده‌اند که ممکن است علت آن جزء متغیرهای این تحقیق نباشد. در گروه سنی ۳۳-۲۸ سال، سابقه ناباروری خانوادگی به احتمال ۸۷/۵٪ در ناباروی اولیه تأثیرگذار بوده است. در سن بالای ۳۳ سال، نسبت خانوادگی زن و شوهر به احتمال ۶۶/۷٪ در ناباروری ثانویه تأثیر داشته است، ولی در گروه زنانی که نامنظمی قاعدگی داشته‌اند، اگر مرد دارای اسپرم کمتر از ۶۸ میلیون در هر میلی‌لیتر بوده به احتمال ۷۶/۵٪ دچار ناباروری اولیه شده‌اند، ولی اگر مقدار اسپرم بالاتر از این عدد بوده، عامل تأثیرگذار در ناباروری اولیه به احتمال ۵۵/۶٪ هیرسوتیسم زن بوده است.

بحث

با توجه به خوشه‌بندی داده‌ها، هریک از گروه بیماران دارای ویژگی‌هایی بوده‌اند که می‌تواند بر روی ناباروری تأثیرگذار باشد. در این مطالعه، میانگین سنی مردان ۳۳-۳۱ سال بود. در مورد نوع ناباروری، در هر دو گروه، بیشتر زوجین دچار ناباروری اولیه بوده و عامل ناباروری در گروه اول که بیشترین تعداد را شامل می‌شد، عامل مردانه عنوان شد که در مطالعات مشابه (۱۹،۱۸) نیز همین نتایج به دست آمد و این احتمال وجود دارد که در این گروه مشکل اصلی، کمبود اسپرم طبیعی مرد بوده است. در گروه دوم احتمال دارد علت ناباروری، عوامل بیماری زنانه باشد، همچنین مردان این گروه دارای شغل با شرایط نشسته بوده‌اند که تأثیر به‌سزایی در کیفیت اسپرم دارد.

در تحقیق حاضر با توجه به عوامل مؤثر در ناباروری براساس مدل CHAID، نمایه توده بدنی زن، مؤثرترین عامل پیش‌بینی شد. همچنین تعداد بسیار کمی از زنانی‌که ناباروری اولیه داشتند دارای BMI بالای ۴۰؛ یعنی چاقی شدید بودند. ۹/۵٪ زنانی‌که ناباروری ثانویه داشتند نیز دارای اضافه وزن و چاقی بودند. براساس مطالعات انجام شده در عربستان سعودی (۲۰) و انگلیس (۲۱)، زنان نابارور به‌طور معنی‌داری چاق‌تر از زنان بارور بوده‌اند.

عامل دوم، سن زن می‌باشد که در مطالعه حاضر گروه سنی زنان بین ۲۹-۲۶ سال بود. در مطالعه مشابه انجام‌شده (۲۲) نیز میانگین سنی زنان ۳۱-۲۷ سال بوده است که به نتایج این مطالعه نزدیک است. تاکنون نتایج بسیاری از مطالعات انجام‌شده در زمینه سن زنان (۲۳) نشان داده است پذیرش رحم (۲۴) و کیفیت تخمک (۲۵)، با افزایش سن کاهش می‌یابد.

عامل سوم تأثیرگذار در ناباروری، بیماری هیرسوتیسم پیش‌بینی شده است. در یک مطالعه انجام‌شده در ساری (۲۶)، شیوع PCOS در بیماران مبتلا به هیرسوتیسم، ۱۸/۷٪ گزارش شد که تقریباً با نتیجه تحقیق حاضر همخوانی داشت. با توجه به جدول خوشه‌بندی، حدود ۴۲/۵٪ زنان مراجعه‌کننده، نامنظمی قاعدگی

داشتند که با توجه به علائم بیماری هیرسوتیسم، نامنظمی قاعدگی در زنان توجیه‌پذیر است. عامل چهارم در تحقیق حاضر، اثر ناباروری خانوادگی بر پیش‌بینی ناباروری تعیین گردید. در یک مطالعه دیگر، اثر ناباروری خانوادگی در مردان نیز مورد تأیید قرار گرفت (۳).

عامل پنجم پیش‌بینی‌شده، گالاکتوره می‌باشد. لذا از آنجایی‌که گالاکتوره یکی از عوارض هیپرپرولاکتینمی است، به تحقیقات انجام‌شده در این زمینه پرداخته شد. هیپرپرولاکتینمی اغلب با اختلال قاعدگی و تخمک‌گذاری همراه است. در بیماران مبتلا به آمنوره و گالاکتوره، هیپرپرولاکتینمی در ۸۰-۶۶٪ موارد یافت می‌شود (۲۸،۲۷). شیوع هیپرپرولاکتینمی در جمعیت نازا در حدود ۲۰٪ می‌باشد (۲۷). براساس معیارهای سازمان بهداشت جهانی، اگر در نتایج نمونه آزمایش اسپرم مردان؛ تعداد اسپرم کمتر از ۲۰×۱۰^{۶} در میلی‌لیتر، مورفولوژی طبیعی کمتر از ۲۰×۱۰^{۶} در میلی‌لیتر و تحرک آن نیز کمتر از ۴۰٪ باشد؛ این مورد به‌عنوان فاکتور مردانه در ناباروری در نظر گرفته می‌شود (۲۹). در تحقیقی با بررسی اثر ارتعاش روی اسپرماتوژنز در مشاغل رانندگان، معدن‌چیان، لکوموتیورانان و غیره، افزایش شیوع اولیگواسپرمی و آزواسپرمی، کاهش حجم منی و درصد اسپرم‌ها با شکل طبیعی مشاهده شد (۳۰). در این تحقیق، ۶٪ مردانی که مشاغل نشسته و ۳٪ مردانی‌که مشاغل ساختمانی داشتند دچار ناباروری ثانویه بودند. در یک تحقیق دیگر با بررسی میزان موفقیت روش IUI در حاملگی، به این نتیجه دست یافتند که با افزایش مدت ناباروری، میزان موفقیت حاملگی کاهش می‌یابد (۳۱). همچنین با افزایش طول مدت ناباروری بیش از ۳ سال به ازای هر سال، میزان کاهش حاملگی به ۹٪ می‌رسد (۳۲). در مطالعه حاضر نیز این نتیجه پیش‌بینی شد که افزایش مدت زمان ناباروری می‌تواند در آن تأثیرگذار باشد.

نتیجه‌گیری

با توجه به نتایج خوشه‌بندی K-means، در خوشه اول به‌نظر می‌رسد بیشترین علت ناباروری، عوامل مردانه بوده و در خوشه دوم، بیشتر عوامل زنانه باعث ناباروری شده است.

با توجه به الگوریتم CHAID، مهم‌ترین عوامل تأثیرگذار در ناباروری به ترتیب شاخص توده بدنی زن، سن زن، بیماری هیرسوتیسم، ناباروری خانوادگی، بیماری گالاکتوره، مقدار اسپرم مرد در هر میلی‌لیتر، مدت ناباروری، سن مرد و نسبت فامیلی زوجین می‌باشد. این مدل، عوامل زنانه را مهم‌تر از عوامل مردانه شناسایی کرده است.

تشکر و قدردانی

بدین وسیله از زحمات سرکار خانم احمری که با راهنمایی‌های خود ما را یاری نمودند و همچنین سرکار خانم معصومه خشک دهان که در جمع‌آوری اطلاعات همراه ما بودند کمال تشکر و قدردانی را داریم.

References:

1. ACECR Center for infertility treatment, Qom Branch. Causes of Infertility. Available From:http://cit-qom.com/fa/index.php/2012-12-21-15-15-1/2013-08-28-07-00-17. Accessed Jul 30, 2014.

2. Widge A. Seeking conception: Experiences of urban Indian woman with in vitro fertilization. Patient Educ Couns 2005;59(3):226-33.

3. Darman KJ, Davise RJ. Distress and marital problems associated with infertility. J Produc and Infant Psychol 1991;5:49-57.

4. Sharghi Sumee N. Comparison of psychosocial status in fertile and infertile women. [MSc Thesis]. School of nursing and midwifery. Beheshti University of Medical Sciences, Tehran; 2003. [Text in Persian]

5. Nene UA, Coyaji K, Apte H. Infertility: A label of choice in the case of sexually dysfunctional couples. Patient Educ Counts 2005;59(3):234-8.

6. Berek JS, Novak E. Brerek & Novak's gynecology. 14th ed. New York: Lippincott Williams & Wilkins; 2006.

7. Speroff LE, Glass RH, Kase NG. Clinical gynecologic endocrinology and infertility. 6th ed. Baltimore, Maryland: Lippincott Williams & Wilkins; 1999. p. 1015-37.

8. Fallahian M. The relationship between employment and fertility problems. Proceedings of the seminar on environmental factors and infertility, Avicenna Research Institute; 2001. p. 12. [Text in Persian]

9. Kimiagar M. Asfrjany F. The role of nutrition in fertility: Review of environmental factors and infertility. Shahid Behshti University of Medical Sciences and Health Services Seminar Papers; 2001. p. 74. [Text in Persian]

10. Hassani Bafrani H, Abedzadeh M, Fruzanfar F,Tabasi Z. Effects of patient age, duration and cause of infertility and number of pre-ovulatory follicles on intrauterine insemination outcomes. J Semnan Univ Med Sci 2010;12(1):59-65. [Full Text in Persian]

11. Fraquhar CM, Lethaby A, Sowter M, Verry J, Baranyai J. An evaluation of risk factors for endometrial hyperplasia in premenopausal women with abnormal menstrual bleeding. Am J Obstet Gynecol 1999;181(3):525-9.

12. Ricci E, Moroni S, Parazzini F, Surace M, Benzi G, Salerio B, Polverino G, La Vecchia C. Risk factors for endometrial hyperplasia: Results from a case-control study. Int J Gynecol Cancer 2002;12(3):257-60.

13. Hornstein MD, Schuts D. Infertility. In: Hillarad PA, Berek JS, Adashi EY. editors. Novakas gynecology. 12th ed. New York: Willams & Wilkins; 1996. p. 915-62.

14. Fehresti Khamseh M, Moghanlofard Y, Moghadam Zarzary Gh. Data Mining an analysis implement managemental. Zanjan: Rozbeh Institute Higher Education; 2014. [Text in Persian]

15. Hoseini M. Developing a predictive model based on the Sarem hospital infertility data. [MSc Thesis]. Tehran: K.N. Toosi University of Technology; 2012. [Text in Persian]

16. Girela JL, Gil D, Johnsson M, Gomez-Torres MJ, De Juan J. Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence. Biol Reprod 2013;88(4):99.

17. Shahrabi J, Zare A. Data Mining with Clementine. Tehran: Amirkabir University; 2013. [Text in Persian]

18. Sedghiani M, Ayatollahi H, Nanbakhsh F, Mojahediyeh A. Success rate of IUI procedure in patients referring to Kowsar IVF Center of Urmia in 2001. J Ardebil Uuiv Med Sci 2006;6(2):142-8. [Full Text in Persian]

19. Yousefi Z, Torabi Zadeh A, Khadem M, Jahanian M. Evaluation of the incidence and outcome of IUI in the Infertility Medical Center, Mashhad University. Med J Mashad Univ Med Sci 2001;44(71):24-19. [Full Text in Persian]

20. Hamilton CJ, Jaroudi KA, Sieck UV. High prevalence of obesity in a saudi infertility population. Ann Saudi Med 1995;15(4):344-6.

21. Pandey S, Bhattacharya S. Impact of obesity on gynecology. Women's Helth (Lond Engl) 2010;6(1):107-17.

22. Abolmasum F, Alizadeh S, Asghari M. Utilizing data mining techniques for investigating factors influencing the failure of intrauterine insemination infertility treatment. J Health Administra 2014;16(54):46-55. [Full Text in Persian]

23. Kang BM, Wu TC. Effect of age on intrauterine insemination with frozen donor sperm. Obstet Gynecol 1996;88(1):93-8.

24. Cano F, Simon C, Remohi J, Peillicer A. Effect of aging on the female reproductive system: Evidence for a role of uterine senescence in the decline in female fecundity. Fertil Steril 1995;64(3):584-9.

25. Abdalla HI, Burton G, Kirkland A, Johnson MR, Leonard T, Brooks AA, et al. Age, pregnancy and miscarriage uterine versus ovarian factors. Hum Reprod 1993;8(9):1512-7.

26. Vasheghani F, Jafari GH, Khan Ahmadi M. Polycystic ovarian disease in the females 15-45 years old referring to endocrine and gynaecology clinic of Imam Khomeini Hospital of Sari Township, 2000-2001. J Mazandaran Univ Med Sci 2002;12(36):52-9. [Full Text in Persian]

27. Behrman SJ, Patton GW, Holtz G. Progress in infertility. 4th ed. Boston: Little, Brown and co; 1986.

28. Minakami H, Abe N, Oka N, Kimura K, Tamura T, Tamada T. Prolactin release in polycystic ovary syndrome. Endoerinal Jpn 1988;35(2):303-10.

29. World Health Organization. WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. 4th ed. New York: Cambridge University Press; 1999.

30. Penkov A, Stainslavov R, Tzvetkov D. Male reproductive function in workers exposed to vibration. Cent Eur J Public Heath 1996;4(3):185-8.

31. Brzechffa PR, Daneshmand S, Buyalos RP. Sequential clomiphene citrate and human menopausal gonadotrophin with intrauterine insemination: The effect of patient age on clinical outcome. Hum Reprod 1998;13(8):2110-4.

32. Collins JA, Rowe TC. Age of female partner is a prognostic factor in prolonged unexplained infertility: A multicenter study. Fertil Steril 1989;52(1):15-20.

Infertility Diagnosis by Data Mining Techniques

Salbi Heydari[1*]; Abolfazl Saeidifar[2]

[1]MSc Student of Social Statistics-Economic, Arak Branch, Islamic Azad University, Qom University of Medical Sciences, Qom, Iran.

[2]Assistant Professor of Mathematics & Statistics, Arak Branch, Islamic Azad University, Arak, Iran.

Corresponding Author:
Salbi Heydari, Arak Branch, Islamic Azad University, Qom University of Medical Sciences, Qom, Iran.

Email:
heydari@muq.ac.ir

Received: 9 Aug, 2014

Accepted: 11 Nov, 2014

Abstract

Background and Objectives: According to wide mass data collection at medical centers and proper use of it in order to diagnosis of a malady needs to relevant tools and medical science for data analyzing. Infertility diagnosis studied by data mining techniques.

Methods: All information had been extract from patient's documents of ACECR Center for Infertility Treatment, Qom Branch; 700 sample were selected among 14,242 cases in 15 years of age, duration of infertility, family connections, infertility, family, job, male, female menstrual cycle type, hirsutism, galactorrhea, amenorrhea, cause of infertility, female body mass index, smoking and semen tests were used. The prediction algorithms C5.0, C & R tree, CHAID and K-means clustering algorithm to determine the optimal number of clusters Davis - Buldian used.

Results: According to the accepted model, the error is less CHAID algorithm, the most important factor in infertility in the female body mass index, age, disease, hirsutism, infertility, family, illness, galactorrhea, the amount of sperm per milliliter, duration of infertility, old man, were consanguineous couples. According to this model, most of the men's wear agents were identified.

Conclusion: In this study, the effect of female infertility factors predicted.

Keywords: Infertility; Data Mining; Algorithms K-means; Decision Trees.