مجله مدیریت اطلاعات سلامت

سال شانزدهم/ شماره اول/ فروردین و اردیبهشت ۱۳۹۸

Published by *Vesnu Publications*

DOI: 10.22122/him.v16i1.3765

پیشنهاد روش درمان ناباروری با استفاده از رده‌بندی ترکیبی و تحلیل داده‌های پرت

رعنا مهدوی[۱]، ثمین فاتحی راویز[۲]، حسین رحمانی[۳]

مقاله پژوهشی

چکیده

مقدمه: با توجه به آمار، امروزه شیوع ناباروری در ایران رو به افزایش است. از طرف دیگر، داده‌کاوی توانسته است با استفاده از داده‌های پزشکی، الگوهای مؤثری را استخراج نماید. پژوهش حاضر با هدف استفاده از داده‌کاوی جهت طراحی سیستمی به منظور پیشنهاد روش درمان ناباروری انجام شد.

روش بررسی: این مطالعه از نوع توصیفی- همبستگی بود که روی اطلاعات ثبت شده از ۵۲۷ زوج نابارور مرکز درمان ناباروری ابن‌سینا تهران انجام گرفت. با بررسی اطلاعات این افراد توسط الگوریتم‌های داده‌کاوی و نرم‌افزار Weka، سیستم PIODEM (Prediction of the best Infertility treatment using Outlier Detection and Ensemble Methods) ارایه شد که شامل سه مرحله بود. ابتدا عوامل مؤثر در انتخاب روش درمان ناباروری با استفاده از تحلیل افتراقی استخراج شد. در مرحله بعد، نمونه‌ها با مقادیر پرت مشخص و ارتباطی بین آن‌ها و انتخاب روش درمان کشف گردید. در نهایت، از رده‌بندهای ترکیبی برای افزایش صحت استفاده شد.

یافته‌ها: سیستم پیشنهادی جهت پیش‌بینی روش درمان، موفق به کشف عوامل مؤثری همچون سن مرد، مدت زمان ناباروری، میزان اسپرم‌های بدون حرکت، کاهش غلظت اسپرم، تعداد کل اسپرم، مورفولوژی، مورفولوژی قسمت میانی اسپرم، اسپرم با حرکت سریع و اسپرم با حرکت کند نوع دوم شد. این سیستم مشخص نمود که پرت بودن مقادیر غلظت اسپرم، توکسوپلاسما IgM (Immunoglobulin M)، هورمون T۳ (Triiodothyronine) و هورمون TPO (Thyroid Peroxidase) در انتخاب روش درمان تأثیرگذار بود. علاوه بر این، استفاده از الگوریتم‌های ترکیبی، معیار F-measure را تا ۷۶ درصد افزایش داد.

نتیجه‌گیری: سیستم PIODEM با استفاده از تحلیل افتراقی و تحلیل داده‌های پرت، قادر به کشف عوامل مؤثر در انتخاب روش درمان می‌باشد. این سیستم با دریافت اطلاعات بیماران به عنوان ورودی، روش درمان را پیشنهاد می‌دهد.

واژه‌های کلیدی: داده‌کاوی؛ ناباروری؛ داده‌های پرت؛ الگوریتم‌های ترکیبی

پیام کلیدی: مطالعه حاضر تلاش کرد تا با بررسی برخی ویژگی‌های زوجین نابارور، روش درمانی مناسب‌تر را به آن‌ها پیشنهاد دهد. یافته‌های به دست آمده می‌تواند جهت پیاده‌سازی سیستم پیشنهاد روش درمان ناباروری در مراکز درمانی مورد استفاده قرار گیرد.

دریافت مقاله: ۱۳۹۷/۸/۱۵ پذیرش مقاله: ۱۳۹۷/۱۲/۲۴ تاریخ انتشار: ۱۳۹۸/۱/۱۵

ارجاع: مهدوی رعنا، فاتحی راویز ثمین، رحمانی حسین. **پیشنهاد روش درمان ناباروری با استفاده از رده‌بندی ترکیبی و تحلیل داده‌های پرت.** مدیریت اطلاعات سلامت ۱۳۹۸؛ ۱۶ (۱): ۱۷-۱۰

مقدمه

آمارها نشان می‌دهد که امروزه ناباروری در ایران به شدت رو به افزایش می‌باشد. ناباروری هنگامی روی می‌دهد که یک زوج پس از گذشت یک سال از زندگی و با وجود داشتن فعالیت جنسی، نتوانند صاحب فرزند شوند (۲، ۱). در ایران نرخ ناباروری ۲۴/۱ درصد گزارش شده است (۳). از طرف دیگر، روش‌های کمک درمانی ناباروری، طولانی و پرهزینه است و گاهی بیماران باید چندین روش درمانی را طی نمایند تا به نتیجه مطلوب برسند (۵، ۴). داده‌کاوی با تجزیه و تحلیل داده‌های موجود زوجین، می‌تواند به شناخت عوامل مؤثر بر ناباروری و پیش‌بینی روش درمانی برای زوجین کمک کند و گام مؤثری جهت کمک به بیماران و پزشکان بردارد (۶).

در پژوهش حاضر، دو روش اصلی درمان ناباروری از جمله روش میکرواینجکشن (۱۰-۷) که شامل تزریق اسپرم داخل تخمک ICSI (Intracytoplasmic Sperm Injection) و روش لقاح درون رحمی IUI (Intrauterine Insemination) (۱۳-۱۱) که شامل گرفتن نمونه اسپرم از مرد و انجام فرایندهای آزمایشگاهی همچون شستن اسپرم و سپس تزریق اسپرم شسته شده به داخل رحم است، مورد بررسی قرار گرفت. با توجه به هزینه و مدت زمان درمان متفاوت این دو روش، استفاده از روش درمانی مناسب برای بیماران اهمیت فراوانی دارد. به دلیل عدم جامعیت مطالعات پیشین، تحقیق حاضر به ارایه سیستم PIODEM (Prediction of the best Infertility treatment using Outlier Detection and Ensemble Methods) پرداخت

مقاله حاصل تحقیق مستقل و بدون حمایت مالی و سازمانی است.

۱- کارشناس ارشد، مهندسی کامپیوتر، گروه نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران (نویسنده طرف مکاتبه)

Email: rana.mahdavii@gmail.com

۲- کارشناس، مهندسی کامپیوتر، گروه هوش مصنوعی و رباتیک، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

۳- استادیار، مهندسی کامپیوتر، گروه نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

که با دریافت اطلاعات زوجین به عنوان ورودی و پردازش آن‌ها، روش درمانی مناسبی را پیشنهاد دهد. به عنوان مثال، Antoniassi و همکاران در پژوهش خود دریافتند که افراد غیر سیگاری به طور معنی‌داری میانگین کیفیت اسپرم بهتر از افراد سیگاری دارند (۱۴). نتایج مطالعه Milewska و همکاران با استفاده از تحلیل افتراقی نشان داد که میانگین سن زنان در گروه موفق در درمان، به طور معنی‌داری کمتر از گروه ناموفق است (۱۵). De Giorgi و همکاران (۱۶) و Hinton (۱۷) در تحقیقات خود با استفاده از تحلیل واریانس برای مقایسه میانگین کیفیت اسپرم در فصول مختلف، به این نتیجه رسیدند که برخی از مؤلفه‌های مربوط به اسپرم مانند میانگین حرکت اسپرم در تابستان بالاتر است. در بررسی دیگری، مظاهری و محسنیان با استفاده از تحلیل واریانس نتیجه‌گیری کردند که ارتباط معنی‌داری میان سطح سلامت روان و وضعیت داشتن فرزند وجود دارد و افراد بارور از سطح سلامت روان بالاتری برخوردار می‌باشند (۱۸). نیلفروشان و همکاران در پژوهش خود با استفاده از تحلیل واریانس، متوجه وجود ارتباط معنی‌دار در میانگین استرس دو گروه پراسترس و کم‌استرس در درمان ناباروری شدند که به تأثیر اضطراب در کاهش نرخ باروری اشاره دارد (۱۹).

نتایج مطالعه عباسی‌هرمزی و همکاران از طریق روش تحلیل واریانس نشان داد مصرف ویتامین D در بهبود حرکت اسپرم مؤثر است (۲۰). تحقیق Chuang و همکاران با استفاده از تحلیل واریانس حاکی از آن بود که افراد با میانگین سطح هورمون FSH (Follicle-Stimulating Hormone) کمتر از ۱۰ واحد بین‌المللی بر میلی‌لیتر و سن کمتر از ۳۵ سال، به طور معنای‌داری نتیجه بهتری از درمان گرفته‌اند (۲۱). پژوهش‌هایی نیز با استفاده از الگوریتم‌های یادگیری ماشین از جمله الگوریتم‌های رده‌بندی و خوشه‌بندی به بررسی ناباروری پرداخته‌اند. به عنوان مثال، Gil و همکاران در مطالعه‌ای با استفاده از درخت تصمیم، به این نتیجه رسیدند که سن، فصل، سابقه برخی از اعمال جراحی و مصرف سیگار، از جمله عوامل تأثیرگذار بر کیفیت اسپرم و مؤلفه‌های آن می‌باشد (۲۲). در تحقیق دیگری، عامری و همکاران با استفاده از خوشه‌بندی K میانگین، دریافتند که بهترین روش درمان نازایی با علت تخمدانی، روش IUI است (۲۳). درمحمدی و همکاران با استفاده از روش خوشه‌بندی K میانگین، به پنج خوشه رسیدند که در هر خوشه بیماران با یکی از عوامل ناباروری تقسیم‌بندی شدند. سن زیاد، مشکل لوله فالوپ و تخمک‌گذاری، اضافه وزن و عوامل مردانه از جمله عوامل مؤثر در ناباروری شناخته شد (۲۴، ۱۴). در پژوهش حفیظ و همکاران، پس از بررسی چندین الگوریتم رده‌بندی پایه و ترکیبی، بالاترین دقت به الگوریتم ترکیبی درخت Random Forest با دقت ۸۳/۹۶ درصد اختصاص یافت. در مطالعه آنان، ویژگی‌های سن زن، تعداد جنین انتقالی و هورمون‌های زنانه به عنوان مهم‌ترین عوامل تأثیرگذار بر نتیجه درمان شناخته شد (۲۵).

تحقیقات پیشین اغلب به بررسی نرخ موفقیت در یک روش خاص درمانی یا عوامل مؤثر بر ابتلا به ناباروری پرداخته‌اند. از طرف دیگر، دو روش درمان ناباروری تفاوت‌های چشمگیری در هزینه و مدت زمان درمان دارند. بنابراین، با توجه به عدم وجود سیستم جامعی جهت پیشنهاد روش درمانی در کارهای پیشین و اهمیت بالای انتخاب روش درمانی مناسب با در نظر گرفتن معیارهای هزینه و زمان، ارایه سیستم جامع PIODEM جهت پیشنهاد روش درمان ناباروری، از اهمیت بالایی برخوردار خواهد بود.

روش بررسی

این پژوهش از نوع توصیفی- همبستگی بود و در آن از داده‌های بیماران نابارور مرکز درمان ناباروری ابن‌سینا تهران استفاده شد. این اطلاعات به ۵۲۷ زوج نابارور، همراه با ۳۱ ویژگی اختصاص داشت. مجموعه داده‌ها شامل افراد تحت درمان روش‌های ICSI و IUI بود. از تعداد ۵۲۷ زوج، ۳۶۲ زوج تحت روش درمانی ICSI و ۱۶۵ زوج تحت روش درمانی IUI قرار گرفته بودند. توزیع داده‌ها و مقادیر کمینه و بیشینه برای هر ویژگی در جدول ۱ آمده است.

مدلی که در مطالعه حاضر جهت ارایه یک سیستم جامع برای پیشنهاد روش درمان به بیماران نابارور ارایه گردید، در شکل ۱ به طور کامل نشان داده شده است. سیستم PIODEM شامل سه مرحله می‌باشد. در مرحله‌ی اول، این سیستم با استفاده از تحلیل افتراقی، معنی‌دار بودن تفاوت میانگین ویژگی‌ها را در بیماران تحت درمان دو روش بررسی کرد. سپس با یافتن رابطه میان پرت بودن نمونه‌ها و روش درمان انتخابی، داده‌های پرت با رویکرد جدیدی تحلیل گردید. در مرحله آخر، با بررسی الگوریتم‌های رده‌بند ترکیبی، سعی در افزایش معیار F-measure در الگوریتم‌ها شد. این معیار یکی از مهم‌ترین معیارها جهت ارزیابی مدل محسوب می‌شود که بر اساس ایجاد تعادلی از دقت و صحت به دست می‌آید و محاسبه آن طبق رابطه ۱ می‌باشد.

رابطه ۱

$$\text{دقت} = \frac{\text{تعداد مثبت‌های درست}}{\text{تعداد مثبت‌های نادرست} + \text{تعداد مثبت‌های درست}}$$

$$\text{فراخوانی} = \frac{\text{تعداد مثبت‌های درست}}{\text{تعداد منفی‌های نادرست} + \text{تعداد مثبت‌های درست}}$$

$$\text{F} - \text{measure} = 2 \times \frac{\text{فراخوانی} \times \text{دقت}}{\text{فراخوانی} + \text{دقت}}$$

مراحل روش پیشنهادی پژوهش جهت ارایه روش درمان ناباروری در شکل ۱ آمده است.

در تحلیل افتراقی، معنی‌داری تفاوت میانگین ویژگی‌ها در دو گروه با دو روش درمانی مختلف محاسبه شد و در نتیجه، ویژگی‌های تأثیرگذار در انتخاب روش درمانی به دست آمد. در تجزیه و تحلیل افتراقی، ابتدا یک متغیر هدف انتخاب می‌گردد و بر اساس مقدار یک ویژگی، نمونه‌ها به دو بخش تقسیم می‌شود. اگر میانگین متغیر هدف و انحراف معیار متغیر برای بخش اول به ترتیب برابر با $\mu 1$، $\mu 2$، $\partial 1$ و $\partial 2$ باشد، رابطه ۲ به یافتن بازه اختلافی دو مجموعه پرداخت که n۱ و n۲ بیان‌کننده تعداد عضو در هر نمونه می‌باشد.

رابطه ۲

$$D = \mu 1 - \mu 2$$

$$SE = \sqrt{\frac{\partial^2}{n1} + \frac{\partial^2}{n2}}$$

$$DiffInterval = D \pm 1.96 * SE$$

Diff Interval، بازه اختلافی متغیر هدف در دو بخش است. اگر این بازه صفر شود، به این معنی است که مقادیر متغیر هدف در این دو بخش اختلاف معنی‌داری ندارد و در غیر این صورت، اختلاف میانگین متغیر هدف در دو بخش معنی‌دارتر خواهد بود.

http://him.mui.ac.ir

جدول ۱: ویژگی‌های زوجین نابارور مورد بررسی

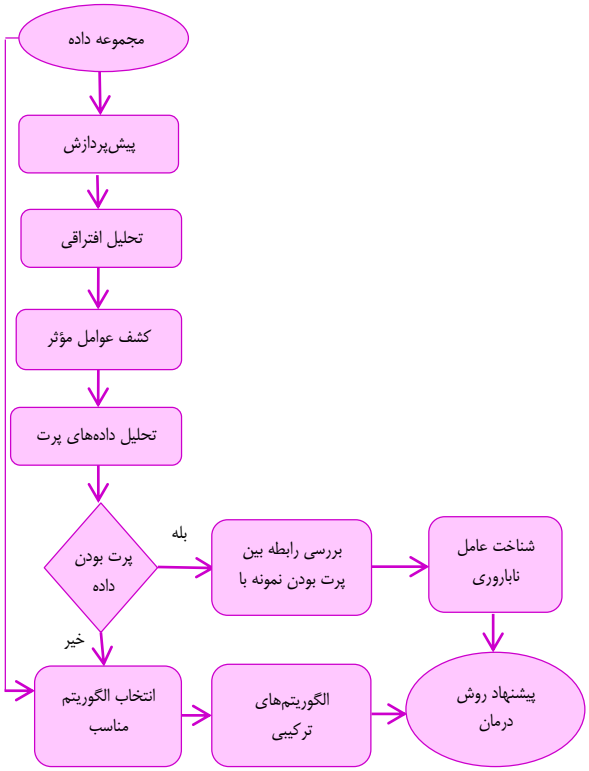
ویژگی‌ها	مقادیر متمایز		میانگین ± انحراف معیار
	کمینه	بیشینه	
سن مرد (سال)	۲۱	۶۰	۳۳/۷ ± ۵/۶
شاخص توده بدنی مرد (کیلوگرم بر مترمربع)	۱۶/۵	۵۰/۱	۲۶/۶ ± ۳/۹
مدت‌زمان ازدواج (سال)	۱	۲۰	۶/۳ ± ۳/۹
طول دوره ناباروری (سال)	۱	۲۰	۴/۵ ± ۳/۷
میزان مصرف سیگار (پاکت- سال)	۰	۲۰	۰/۴۸ ± ۲/۱
هورمون تستوسترون (پیکوگرم بر میلی‌لیتر)	۰/۱۱	۱۳/۶	
هورمون LH (واحد بین‌المللی بر میلی‌لیتر)	۰/۱	۴۳/۴	۷/۰ ± ۶/۴
هورمون FSH (واحد بین‌المللی بر میلی‌لیتر)	۰/۲	۶۴/۳	۱۱/۰۱ ± ۱۲/۱
غلظت اسپرم (تعداد اسپرم در میلی‌لیتر)	۰	۴۵۰	۴۹/۹ ± ۵۲/۱
تعداد کل اسپرم	۰	۱۰۳۹/۵	۱۵۴/۸ ± ۱/۷
حجم اسپرم (میلی‌لیتر)	۰/۱	۱۱/۰	۳/۲۴ ± ۱/۷
غیر طبیعی بودن مورفولوژی اسپرم (درصد)	۷۰/۰	۱۰۰	۹۳/۵ ± ۵/۱۵
غیر طبیعی بودن مورفولوژی سر اسپرم (درصد)	۲۱	۹۷	۷۳/۲ ± ۹/۳
غیر طبیعی بودن مورفولوژی قسمت میانی (درصد)	۰	۱۴	۳/۸ ± ۲/۱
غیر طبیعی بودن مورفولوژی دم اسپرم (درصد)	۲	۷۳	۸/۷ ± ۵/۹
حرکت اسپرم با سرعت بالا (درصد)	۰	۷۰	۱۰/۲ ± ۱۰/۶
حرکت اسپرم با سرعت نوع دوم (درصد)	۰	۶۵	۲۱/۵ ± ۱۰/۱
حرکت اسپرم با سرعت نوع سوم (درصد)	۰	۷۵	۱۲/۷ ± ۷/۳
اسپرم بدون حرکت (درصد)	۱۵	۱۰۰	۵۵/۶ ± ۱۶/۶
سن زن (سال)	۱۷	۴۱	۲۸/۹ ± ۴/۶
شاخص توده بدنی زن (کیلوگرم بر مترمربع)	۱۴/۸	۴۰	۲۵/۱ ± ۳/۸
توکسوپلاسما IgG (واحد بین‌المللی بر میلی‌لیتر)	۰/۶	۵۰۰	۳۲/۷ ± ۸۰/۱
توکسوپلاسما IgM (واحد بین‌المللی بر میلی‌لیتر)	۰/۱	۴۸/۹	۳/۲ ± ۳/۵
تیروگلوبین (واحد بین‌المللی بر میلی‌لیتر)	۰/۲	۴۰۰۰	۱۲۴/۶ ± ۳۹۴/۷
هورمون TPO (واحد بین‌المللی بر میلی‌لیتر)	۰/۳	۱۲۲۱	۴۰/۱ ± ۱۱۰/۵
هورمون TSH (واحد بین‌المللی بر لیتر)	۰	۱۷/۵	۲/۳ ± ۱/۵
هورمون FSH (واحد بین‌المللی بر لیتر)	۰/۲	۳۷	۷/۱ ± ۳/۴
هورمون LH (واحد بین‌المللی بر لیتر)	۰/۰۷	۱۹۵/۲	۶/۶ ± ۱۱/۷
هورمون پرولاکتین (واحد بین‌المللی بر لیتر)	۰/۹	۱۴۴۸	۲۳۴/۰ ± ۲۵۸/۸
هورمون T۳ (واحد بین‌المللی بر میلی‌لیتر)	۱	۲۸۰	۸/۰ ± ۲۸/۱
هورمون T۴ (میکروگرم بر دسی‌لیتر)	۱	۱۳۶	۲/۵ ± ۸/۳

LH: Luteinizing Hormone; FSH: Follicle Stimulating Hormone; IgG: Immunoglobulin G; TPO: Thyroid Peroxidase; TSH: Thyroid Stimulating Hormone

بررسی دیگر صورت گرفته، تحلیل و محاسبه داده‌های پرت بود. در تحقیق حاضر نمونه‌هایی که حداقل در یک ویژگی مقدار پرت داشتند، به عنوان نمونه‌های پرت محسوب شدند. مقادیری که خارج از بازه حد بالا و پایین محاسبه شده از طریق نمودار جعبه‌ای بودند، مقادیر پرت به شمار می‌آیند (۲۶). به نظر می‌رسد که پزشک یا سیستم داده‌کاوی با مشاهده مقادیر پرت در ویژگی‌های بیماران، عامل ناباروری را کشف می‌کند و قادر است به صورت دقیق‌تر روش درمان مناسب را بر اساس مقادیر آن ویژگی پیشنهاد نماید. جهت اثبات و ارزیابی فرضیه رابطه میان پرت بودن هر نمونه با انتخاب روش درمان، آزمون χ^2 مورد استفاده قرار گرفت.

یکی از راهکارهای افزایش دقت رده‌بندها، استفاده از الگوریتم‌های ترکیبی است و از آن می‌توان جهت یافتن عوامل مؤثر بهره برد. الگوریتم‌های رده‌بندی Bagging و Boosting، از متداول‌ترین الگوریتم‌های ترکیبی به شمار می‌روند.

http://him.mui.ac.ir



شکل ۱: مراحل روش پیشنهادی جهت ارایه روش درمان مناسب

الگوریتم‌های ترکیبی، الگوریتم‌های پایه ضعیف را به الگوریتمی قوی تبدیل می‌کنند و به دو دسته همگون و ناهمگون تقسیم می‌شوند. در پژوهش حاضر، ابتدا الگوریتم‌های ترکیبی همگون Bagging و AdaBoost روی مجموعه داده‌ها اعمال گردید و سپس جهت افزایش دقت الگوریتم‌ها، از روش ناهمگون استفاده شد. برای عملکرد بهتر الگوریتم‌ها و افزایش دقت در خروجی، انتخاب الگوریتم‌های پایه جهت ترکیب و ساخت الگوریتم ترکیبی باید دو شرط اصلی را داشته باشد؛ اول این که الگوریتم‌های پردقت انتخاب گردد و دوم این که ساختار الگوریتم‌ها متنوع باشد. برای بررسی و اعمال الگوریتم‌های پایه در مطالعه حاضر، از نرم‌افزار Weka (۲۸، ۲۷) و برای ارزیابی از 10-fold Cross Validation استفاده شد. این الگوریتم‌ها در پنج گروه « الگوریتم‌های مبتنی بر Bayesian، الگوریتم‌های مبتنی بر تابع، الگوریتم‌های تنبل، الگوریتم‌های مبتنی بر قانون و الگوریتم‌های درخت تصمیم» قرار می‌گیرند. یکی از روش‌های ترکیب، روش اکثریت آرا به کمک وزن است. زمانی که الگوریتم‌های پایه دقت‌های متفاوتی داشته باشند، به الگوریتم‌های قوی‌تر با دقت بالا، وزن بیشتری اختصاص می‌یابد و نتایج نهایی بر اساس رأی اکثریت داده می‌شود. در این روش، وزن هر کدام از الگوریتم‌ها بر اساس رابطه ۳ محاسبه می‌گردد.

$$Wi = Log \frac{ai}{1-ai}$$ رابطه ۳

لازم به ذکر است که کلیه ملاحظات اخلاقی در زمینه حفظ محرمانگی اطلاعات شخصی بیماران در تحقیق حاضر رعایت شده است.

یافته‌ها

نتایج تحلیل افتراقی در جدول ۲ ارایه شده است. بر این اساس، با افزایش مقادیر ویژگی‌های سن مرد، مدت زمان ناباروری و میزان اسپرم‌های بدون حرکت، سیستم PIODEM روش درمان ICSI را به عنوان خروجی پیشنهاد می‌کند و در صورت افزایش غلظت اسپرم، تعداد کل اسپرم در نمونه اخذ شده، مورفولوژی، مورفولوژی قسمت میانی اسپرم، اسپرم با حرکت سریع و اسپرم با حرکت کند نوع دوم نیز روش IUI پیشنهاد خواهد شد.

جدول ۲: نتایج تحلیل افتراقی در انتخاب روش درمان

انتخاب روش‌های درمان IUI/ICSI	
بازه	تحلیل افتراقی
۰/۶۲ ،۲/۵۱	سن مرد
۰/۸۷ ،۲/۱۰	مدت زمان ناباروری
-۲۶/۱۱ ،-۴۶/۳۰	غلظت اسپرم
-۸۰/۲۷ ،-۱۴۷/۱	تعداد اسپرم
۴/۴۰ ،۶/۴۶	مورفولوژی قسمت میانی اسپرم
۰/۱۲ ،۱/۰۱	مورفولوژی
-۳/۷۹ ،-۸/۴۵	درصد اسپرم با حرکت پرسرعت
-۵/۱۲ ،-۸/۸۲	درصد اسپرم با حرکت نسبتاً کند
۹/۱۹ ،۱۵/۲۷	درصد اسپرم بدون حرکت

IUI/ICSI: Intrauterine Insemination/Intracytoplasmic Sperm Injection

در تحلیل داده‌های پرت، ارتباط میان پرت بودن هر نمونه و انتخاب روش درمان با استفاده از آزمون χ^2 بررسی گردید که در جدول ۳ نشان داده شده است. مقدار P، میزان احتمال عدم ارتباط ویژگی مورد نظر با انتخاب روش درمان می‌باشد. هرچه این مقدار کمتر باشد، ویژگی مورد نظر ارتباط بیشتری با انتخاب روش درمان دارد.

جدول ۳: نتایج آزمون χ^2 برای بررسی رابطه میان نمونه‌های پرت و انتخاب روش درمان

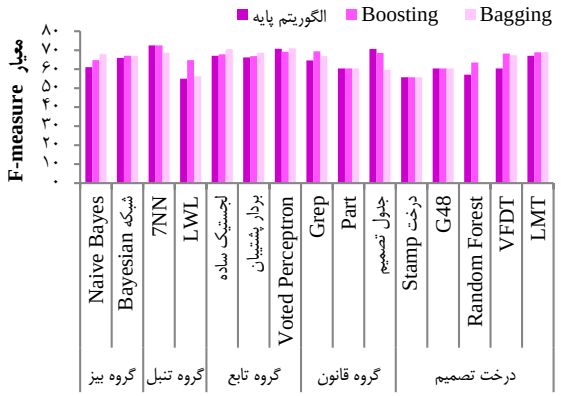
ویژگی	آزمون χ^2	مقدار P
مورفولوژی	۹/۸۷	۰/۰۰۲
غلظت اسپرم	۵/۷۲	۰/۰۱۰
توکسوپلاسما IgM	۵/۶۶	۰/۰۱۰
هورمون T3 زنان	۶/۱۰	۰/۰۱۰
هورمون TPO زنان	۵/۸۹	۰/۰۱۰
تیروگلوبین	۴/۵۸	۰/۰۳۰

IgM: Immunoglobulin M; TPO: Thyroid Peroxidase

یافته‌های جدول ۳ نشان داد، نمونه‌هایی که در آن‌ها ویژگی‌های مورفولوژی، توکسوپلاسما IgM، هورمون‌های T3، TPO زنان و تیروگلوبین مقادیر پرت داشته باشد، بیشتر از دیگر نمونه‌ها تحت درمان روش ICSI بودند. افرادی که در ویژگی غلظت اسپرم مقداری پرت داشتند، بیشتر تحت روش درمان IUI قرار گرفتند.

تحلیل داده‌های پرت در شناخت عوامل تأثیرگذار بر انتخاب روش مؤثر است، اما برای پیش‌بینی روش درمان در مرحله بعد، از رده‌بندها استفاده گردید.

در این مرحله، الگوریتم‌های پنج گروه مذکور، اجرا و الگوریتم‌های Bagging و Boosting بر هر یک از آن‌ها اعمال شد. شکل ۲ بیان‌کننده معیار F-measure برای هر یک از الگوریتم‌ها می‌باشد. بر این اساس، بیشترین معیار F-measure مربوط به الگوریتم ۷ نزدیک‌ترین همسایه و برابر با ۷۲/۵ درصد بود.



شکل ۲: ارزیابی الگوریتم‌های پایه و اعمال الگوریتم‌های Bagging و Boosting روی هر یک

7NN: 7 Nearest Neighbors; LMT: Logistic Model Tree; LWL: Locally Weighted Learning; VFDT: Very Fast Decision Tree

در مرحله بعد، از میان الگوریتم‌های پایه هر گروه، الگوریتم با بالاترین F-measure انتخاب و از آن جهت ترکیب الگوریتم استفاده گردید (شکل ۳). وزن‌دهی الگوریتم‌ها با استفاده از رابطه ۳ محاسبه و برای ترکیب الگوریتم‌ها از روش اکثریت آرا استفاده شد و F-measure به ۷۶ درصد افزایش یافت.

شکل ۳. نتایج الگوریتم‌های ترکیبی ناهمگون در انتخاب روش درمان

7NN: 7 Nearest Neighbors; LMT: Logistic Model Tree

برای ارزیابی الگوریتم ترکیبی حاصل، می‌توان از نمودار ROC (Receiver Operating Characteristic) استفاده نمود. میزان ارزشمند بودن هر الگوریتم با سطح نمودار ROC آن رابطه مستقیمی دارد. مساحت زیر نمودار الگوریتم‌های پایه و الگوریتم ترکیبی در جدول ۴ نشان داده شده است. بر این اساس، می‌توان نتیجه گرفت که استفاده از الگوریتم ترکیبی، رده‌بندی را بهبود بخشیده است.

جدول ۴: سطح زیر نمودار Receiver Operating Characteristic (ROC) برای الگوریتم‌های ترکیبی ناهمگون در انتخاب روش درمان

الگوریتم رده‌بندی	سطح زیر نمودار ROC
Naive Bayes	۰/۷۳۰۶
vNN	۰/۷۳۷۳
Voted Perceptron	۰/۷۳۰۶
جدول تصمیم	۰/۷۰۲۱
LMT	۰/۷۲۹۷
الگوریتم ترکیبی	۰/۷۶۵۴

ROC: Receiver Operating Characteristic; 7NN: 7 Nearest Neighbors; LMT: Logistic Model Tree

بحث

مهم‌ترین یافته پژوهش حاضر، ارایه سیستم جامع PIODEM جهت پیشنهاد روش درمان برای بیماران نابارور می‌باشد. این سیستم در مرحله اول PIODEM موفق به شناسایی عوامل مؤثر بر انتخاب روش درمان می‌شود. بر این اساس، افزایش سن مرد، مدت زمان ناباروری، میزان اسپرم‌های بدون حرکت، کاهش غلظت اسپرم، تعداد کل اسپرم در نمونه اخذ شده، مورفولوژی، مورفولوژی قسمت میانی اسپرم، اسپرم با حرکت سریع و اسپرم با حرکت کند، منجر به پیشنهاد روش درمان ICSI شد. در صورت کاهش خصیصه‌های فوق، PIODEM روش درمان IUI را پیشنهاد خواهد کرد. همچنین، نتایج سیستم PIODEM نشان می‌دهد که اگر نمونه‌ها در ویژگی‌های مورفولوژی، غلظت اسپرم، توکسوپلاسما IgM، هورمون T۳ و هورمون TPO مقادیر پرت داشته باشند، پزشک و یا سیستم داده‌کاوی می‌تواند به طور مشخص روش درمان ناباروری را انتخاب نماید.

لازم به ذکر است که با توجه به مطالعات پیشین و با توجه به نظر متخصص ناباروری در تشخیص روش درمان ناباروری، کمتر به اهمیت ویژگی‌هایی مانند سن مرد، مورفولوژی قسمت میانی اسپرم و هورمون زنانه TPO پرداخته شده است؛ در حالی که نتایج تحقیق حاضر حاکی از آن است که این ویژگی‌ها مهم به شمار می‌روند و در انتخاب روش درمان باید به مقادیرشان توجه شود که این نتیجه از نظر متخصصان دامنه ارزشمند و مورد تأیید می‌باشد. از چالش‌های موجود در مطالعه حاضر می‌توان به نقص داده در بسیاری از ویژگی‌ها اشاره کرد. به عنوان مثال، میزان مصرف سیگار در بیماران به طور ناقص ثبت گردید. همچنین، عدم ثبت اطلاعات به صورت دیجیتال، جمع‌آوری آن‌ها را با مشکل مواجه کرد. از طرف دیگر، عدم دسترسی به اطلاعات تعداد بیشتر بیماران، از دیگر مشکلاتی بود که در صورت تعدیل این محدودیت، به طور قطع پژوهش به نتایج بهتری دست می‌یافت.

نتیجه‌گیری

سیستم PIODEM با دریافت ویژگی‌های بیماران نابارور به عنوان ورودی و همچنین، با استفاده از تحلیل داده‌های پرت و بهره‌گیری از الگوریتم‌های ترکیبی، روش درمان ناباروری را به عنوان خروجی به زوجین نابارور

پیشنهاد می‌نماید.

پیشنهادها

با توجه به نیاز در حوزه ناباروری، می‌توان توسعه نرم‌افزاری نتایج تحقیق حاضر را به صورت برنامه‌های کاربردی توصیه نمود. همچنین، برقراری شبکه‌های اجتماعی برای پیشنهاد روش درمانی، اطلاع‌رسانی به بیماران و جمع‌آوری اطلاعات می‌تواند مفید باشد.

تشکر و قدردانی

بدین وسیله از کلیه استادانی که در انجام مطالعه حاضر همکاری نمودند، تشکر و قدردانی به عمل می‌آید.

تضاد منافع

در انجام مطالعه حاضر، نویسندگان هیچ‌گونه تعارض منافعی نداشته‌اند.

References

1. Turchi P. Prevalence, definition, and classification of infertility. In: Cavallini G, Beretta G, editors. Clinical management of male infertility. Cham, Switzerland: Springer International Publishing; 2015. p. 5-11.
2. Gurunath S, Pandian Z, Anderson RA, Bhattacharya S. Defining infertility--a systematic review of prevalence studies. Hum Reprod Update 2011; 17(5): 575-88.
3. Kazemijaliseh H, Ramezani TF, Behboudi-Gandevani S, Hosseinpanah F, Khalili D, Azizi F. The prevalence and causes of primary infertility in Iran: A population-based study. Glob J Health Sci 2015; 7(6): 226-32.
4. Klitzman R. How much is a child worth? Providers' and patients' views and responses concerning ethical and policy challenges in paying for ART. PLoS One 2017; 12(2): e0171939.
5. Bhide A. Fertility treatment: Getting stressed about stress. Acta Obstet Gynecol Scand 2018; 97(3): 233-4.
6. Durairaj M, Ramasamy N. Intelligent prediction methods and techniques using disease diagnosis in medical database: A review. International Journal of Control Theory and Applications 2015; 8(5): 2153-60.
7. Suzuki T, Perry ACF. Intracytoplasmic sperm injection (ICSI): Applications and insights. In: Palermo GD, Sills ES, editors. Intracytoplasmic sperm injection: Indications, techniques and applications. Cham, Switzerland: Springer International Publishing; 2018. p. 169-81.
8. Vaegter KK, Lakic TG, Olovsson M, Berglund L, Brodin T, Holte J. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. Fertil Steril 2017; 107(3): 641-8.
9. Mckinnon AO, Trounson AO, Silber SJ. Intracytoplasmic sperm injection. In: Samper JC, Pycock JF, Mckinnon AO, editors. Current therapy in equine reproduction. Saint Louis, CA: W.B. Saunders; 2007. p. 296-307.
10. Meijerink AM, Cissen M, Mochtar MH, Fleischer K, Thoonen I, de Melker AA, et al. Prediction model for live birth in ICSI using testicular extracted sperm. Hum Reprod 2016; 31(9): 1942-51.
11. Ombelet W, Dhont N, Thijssen A, Bosmans E, Kruger T. Semen quality and prediction of IUI success in male subfertility: A systematic review. Reprod Biomed Online 2014; 28(3): 300-9.
12. Monraisin O, Chansel-Debordeaux L, Chiron A, Floret S, Cens S, Bourrinet S, et al. Evaluation of intrauterine insemination practices: a 1-year prospective study in seven French assisted reproduction technology centers. Fertil Steril 2016; 105(6): 1589-93.
13. Mookim PG, Ellis RP, Kahn-Lang A. Infertility treatment, ART and IUI procedures and delivery. Outcomes: How important is selection? Boston MA: Boston University; 2010. [Unpulished].
14. Antoniassi MP, Intasqui P, Camargo M, Zylbersztejn DS, Carvalho VM, Cardozo KH, et al. Analysis of the functional aspects and seminal plasma proteomic profile of sperm from smokers. BJU Int 2016; 118(5): 814-22.
15. Milewska AJ, Jankowska D, Cwalina U, Citko D, Wiesak T, Acacio B, et al. Significance of discriminant analysis in prediction of pregnancy in IVF treatment. Studies in Logic, Grammar and Rhetoric 2015; 43(1): 7-20.
16. De Giorgi A, Volpi R, Tiseo R, Pala M, Manfredini R, Fabbian F. Seasonal variation of human semen parameters: A retrospective study in Italy. Chronobiol Int 2015; 32(5): 711-6.
17. Hinton PR. Statistics explained. London, UK: Routledge; 2014. p. 125-31.
18. Mazaheri M, Mohsenian R. Comparison of mental health ratings of fertile and infertile couples. Zahedan J Res Med Sci 2012; 14(1): 72-5. [In Persian].
19. Nilforooshan P, Ahmadi SA, Abedi MR, Ahmadi SM. Attitude towards infertility and its relation to depression and anxiety in infertile couples. J Reprod Fertil 2006; 6(5): 546-53.
20. Abbasihormozi S, Kouhkan A, Alizadeh AR, Shahverdi AH, Nasr-Esfahani MH, Sadighi Gilani MA, et al. Association of vitamin D status with semen quality and reproductive hormones in Iranian subfertile men. Andrology 2017; 5(1): 113-8.
21. Chuang CC, Chen CD, Chao KH, Chen SU, Ho HN, Yang YS. Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. Fertil Steril 2003; 79(1): 63-8.
22. Gil D, Girela JL, De Juan J, Gomez-Torres MJ, Johnsson M. Predicting seminal quality with artificial intelligence methods. Expert Syst Appl 2012; 39(16): 12564-73.
23. Ameri H, Alizadeh S, Hadizadeh M. Assessing the effects of infertility treatment drugs using clustering algorithms and data mining techniques. J Mazandaran Univ Med Sci 2014; 24(114): 26-35. [In Persian].

24. Dormohammadi S, Alizadeh S, Asghari M, Shami M. Proposing a prediction model for diagnosing causes of infertility by data mining algorithms. J Health Adm 2014; 17(57): 46-57. [In Persian].
25. Hafiz P, Nematollahi M, Boostani R, Namavar JB. Predicting implantation outcome of in vitro fertilization and intracytoplasmic sperm injection using data mining techniques. Int J Fertil Steril 2017; 11(3): 184-90.
26. Han J, Pei J, Kamber M. Data mining: Concepts and techniques. Burlington, MA: Morgan Kaufmann; 2011.
27. Frank E, Hall MA, Witten IH. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016 [Online]. [cited 2016]; Available from: URL: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf
28. Desai A, Rai S. Analysis of Machine Learning Algorithms using Weka. Proceedings of the International Conference and Workshop on Recent Trends in Technology, (TCET) 2012. International Journal of Computer Applications 2012; 27-32.

Health Information Management

Published by ***Vesnu Publications***

Vol 16/ No 1/ April & May 2019

DOI: 10.22122/him.v16i1.3765

Suggesting the Infertility Treatment Method Using Ensemble Methods and Outlier Analysis

Raana Mahdavi[1], Samin Fatehi-Raviz[2], Hossein Rahmani[3]

Original Article

Abstract

Introduction: In recent years, the infertility ratio in young couples has been increased a lot in Iran. From the other side, it has been shown that data mining techniques are capable of extracting novel patterns from medical data. In this study, we proposed a comprehensive system called Prediction of the best Infertility treatment using Outlier Detection and Ensemble Methods (PIODEM) for predicting of the best infertility treatment method for infertile couples.

Methods: This descriptive-correlation study used the information of 527 infertile couples, which collected from Avicenna specialized infertility center, Tehran, Iran. PIODEM consists of three steps: First, PIODEM uses the discriminant analysis to find effective factors for choosing the best infertility treatment. Second, PIODEM detects the outlier samples, and applies a correlation between these samples and the choice of treatment method. Third, it uses ensemble methods to increase the precision of classifiers.

Results: The PIODEM system succeeded in discovering affective factors such as male-partner's age, infertility duration, immotile sperm, decreasing of sperm concentration decrease, total sperm count, morphology, sperm motility, sperm with rapid progressive-a motility, and sperm with slow progressive-b motility. Additionally, PIODEM indicates that if one of four features of sperm concentration, toxoplasma immunoglobulin M (IgM), triiodothyronine (T3) hormone, and thyroid peroxidase (TPO) was an outlier, then the prediction of treatment would be more accurate. Finally, using ensemble methods increased the F-measure of PIODEM system by up to 76%.

Conclusion: The PIODEM system is able to discover effective factors in the choice of treatment method, using differential analysis and analysis of pert data. This system offers patient information as input for the treatment method.

Keywords: Data Mining; Infertility; Outlier Analysis; Ensemble Algorithms

Received: 06 Nov., 2018 Accepted: 15 Mar., 2019 Published: 04 Apr., 2019

Citation: Mahdavi R, Fatehi-Raviz S, Rahmani H. **Suggesting the Infertility Treatment Method Using Ensemble Methods and Outlier Analysisthe Infertility Treatment Method Using Ensemble Methods and Outlier Analysis.** Health Inf Manage 2019; 16(1): 10-7

Article resulted from an independent research without financial support.

1- MSc, Computer Engineering, Department of Software, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran (Corresponding Author) Email: rana.mahdavii@gmail.com

2- BSc, Computer Engineering, Department of Artificial Intelligence and Robotics, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

3- Assistant Professor, Computer Engineering, Department of Software, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran