SJRM SAREM JOURNAL OF REPRODUCTIVE MEDICINE

Sarem Journal of Reproductive Medicine. 2017;2(2):105-112

Analysis of the Factors on Intrauterine Insemination (IUI) Results by Clustering

ARTICLE INFO

Article Type
Original research

Authors
Alizadeh S.[1] *PhD*,
Asghari M.* *MSc*,
Hosseini M.K.[1] *MSc*

How to cite this article
Alizadeh S, Asghari M, Hosseini M K. Analysis of the Factors on Intrauterine Insemination (IUI) Results by Clustering. Sarem Journal of Reproductive Medicine. 2017;1(3):105-112.

ABSTRACT

Aims Intra Uterine Insemination (IUI) is a medically-assisted reproduction technique (ART) enables infertile couples to achieve the successful pregnancy. Given the unpredictability of such techniques, many investigations have been done on the factors affecting the techniques. Data mining is one of the main tools that can help researchers to evaluate the factors. Data mining utilize the statistical methods along with the artificial intelligence (AI) to help different sciences including infertility science and research for interpreting the results and analyzes of data appropriately and extracting the hidden patterns and knowledge in the data. The objective of this study was to analyze the factors affecting IUI results by clustering.

Materials & Methods The IUI data were clustered utilizing the K-means)a clustering method in data mining). Davise-Buldian index was used to calculate the best number of clusters. The similar individuals were included in the same cluster and the success rates in those clusters were also measured.

Findings Some of the characteristics of individuals such as age, body mass index (BMI), type of infertility, the cause of infertility and etc. were effective factors on IUI success rate.

Conclusion Factors such as age, BMI, type of infertility, the cause of infertility and etc. can determine the success rate of the IUI method.

Keywords Infertility; Insemination, Artificial; Clustering; Data Mining

*Information Technology Department, Computer Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran
[1]Information Technology Department, Computer Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran

Correspondence
Address: -
Phone: -
Fax: -
mohsen.asghari@gmail.com

Article History
Received: February 26, 2016
Accepted: June 21, 2016
ePublished: August 15, 2017

CITATION LINKS

[1] Data Mining Curriculum: A Proposal (Version 1.0) [2] Diabetes-like renal glomerular disease in Fanconi-Bickel syndrome [3] Data Mining: Concepts and Techniques [4] Infertility Guideline [5] Fertility outcome after IVF and related factors [6] Bayesian classification for the selection of in vitro human embryos using morphological and clinical data [7] The relationship between number of transferred embryos and pregnancy rate in ART cycles [8] Data Mining and Knowledge Discovery [9] Sodium butyrate activates erythroid-specific 5-aminolevulinate synthase gene through Sp1 elements at its promoter [10] A comparison in cluster validation techniques [11] Data Mining: Practical Machine Learning Tools and Techniques [12] Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes [13] Case-based reasoning in IVF: Prediction and knowledge mining [14] Factors predicting the cumulative outcome of IVF/ICSI treatment: A multivariable analysis of 2450 patients

Copyright© 2017, ASP Ins. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.

تحلیل عوامل موثر بر نتایج روش انتقال داخل‌رحمی اسپرم با استفاده از خوشه‌بندی

سمیه علیزاده PhD
گروه فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران
محسن اصغری* MSc
گروه فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران
محمدکاظم حسینی MSc
گروه فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

چکیده
اهداف: انتقال داخل‌رحمی اسپرم (IUI) یکی از روش‌هایی است که امکان باردارشدن به زوج‌های نابارور را می‌دهد. غیرقابل پیش‌بینی‌بودن این روش‌ها موجب شده است تا بررسی‌های زیادی روی عوامل موثر بر آنها انجام شود. در همین راستا، علم داده‌کاوی با استفاده از مفاهیم جدید هوش مصنوعی و علوم آماری به کمک علوم مختلف از جمله علم ناباروری آمده است تا به تفسیر نتایج و تحلیل‌های مناسب داده‌ها کمک کند و بتواند الگوها و دانش‌های نهفته در داده‌ها را استخراج کند. این مطالعه با هدف تحلیل عوامل موثر بر نتایج روش انتقال داخل‌رحمی اسپرم با استفاده از خوشه‌بندی انجام شد.
مواد و روش‌ها: با استفاده از تکنیک K-means (روش توصیفی در داده‌کاوی)، داده‌ها به چندین خوشه تقسیم شدند و از تکنیک دیویس- بولدین برای محاسبه تعداد بهینه خوشه‌ها استفاده شد. در این توصیف‌ها افرادی که به هم شبیه بودند در یک خوشه قرار گرفتند و میزان موفقیت در این خوشه‌ها اندازه‌گیری شد.
یافته‌ها: برخی از ویژگی‌های افراد مانند سن، توده حجمی بدن (BMI)، نوع ناباروری، علت ناباروری و غیره در میزان موفقیت روش IUI می‌تواند تاثیرگذار و تعیین‌کننده باشد.
نتیجه‌گیری: عواملی مانند سن، BMI، نوع ناباروری، علت ناباروری و غیره می‌توانند میزان موفقیت روش IUI را تعیین کنند.
کلیدواژه‌ها: ناباروری، داده‌کاوی، خوشه‌بندی، K-means، روش IUI

تاریخ دریافت: ۱۳۹۴/۱۲/۰۷
تاریخ پذیرش: ۱۳۹۵/۰۴/۰۱
*نویسنده مسئول: mohsen.asghari@gmail.com

مقدمه

امروزه پیشرفت روزافزون علوم مختلف در زمینه فناوری اطلاعات، موضوع مهمی در حیطه علوم است. یکی از رشته‌هایی که پیشرفت‌های آشکاری دارد، علم پزشکی است. تکنیک‌های داده‌کاوی از جمله روش‌های نوین در علم پزشکی هستند. داده‌کاوی، یک حوزه میان‌رشته‌ای با رشد سریع است که حوزه‌های مختلفی مانند پایگاه داده، آمار، یادگیری ماشین و سایر زمینه‌های مرتبط را ترکیب می‌کند تا اطلاعات و دانش ارزشمند نهفته در حجم بزرگی از داده‌ها را استخراج کند[1]. داده‌کاوی، اکتشاف و تحلیل حجم زیادی از داده‌ها برای کشف الگوها و قواعد معنی‌دار است. فرآیند داده‌کاوی گاهی "کشف دانش" نیز نامیده می‌شود[2]. روش‌های اصلی داده‌کاوی به دو دسته توصیفی و پیش‌بینی تقسیم می‌شوند. هدف از توصیف، یافتن الگوهای قابل توصیف برای انسان، در مورد داده‌ها است[3]. روش پیش‌بینی به‌منظور پیش‌بینی رفتارهای آینده آنها استفاده می‌شود. از رایج‌ترین الگوریتم‌های توصیفی استفاده‌شده پیرامون داده‌های مختلف، تحلیل خوشه‌بندی است. در این پژوهش نیز از این الگوریتم برای تحلیل داده‌ها استفاده شده است. براساس تعاریف سازمان بهداشت جهانی، نازایی یا ناباروری به معنی عدم وقوع حاملگی، یک سال پس از ازدواج یا از زمانی که زوجین، تصمیم به بچه‌دارشدن می‌گیرند (بدون استفاده از روش‌های پیشگیری) است[4]. به‌طور کلی علل مختلف ناباروری را عوامل مردانه، زنانه و نامعین تشکیل می‌دهند. از درمان‌های رایج در ناباروری، تکنیک‌های کمک‌باروری (ART) هستند که نقش موثری در درمان ناباروری با علل گوناگون دارند. از جمله تکنیک‌های کمک‌باروری رایج می‌توان به تزریق اسپرم به داخل سیتوپلاسم تخمک (ICSI)، انتقال گامت درون فالوپ (GIFT)، تلقیح آزمایشگاهی (IVF)، تلقیح درون رحمی (IUI) و انتقال تخم درون فالوپ (ZIFT) اشاره کرد[4].

داده‌کاوی در ناباروری: با استفاده از تکنیک‌های داده‌کاوی می‌توان در تشخیص و درمان بیماری‌ها به پزشکان کمک کرد. از جمله زمینه‌های پزشکی که می‌توان از تکنیک‌های داده‌کاوی در آن استفاده کرد، ناباروری است. می‌توان از این تکنیک‌ها برای تشخیص بهتر و سریع‌تر علت ناباروری و همچنین پیشنهاد درمان بهتر با احتمال موفقیت بالاتر استفاده کرد. از روش‌های مختلف داده‌کاوی می‌توان برای تشخیص و درمان سریع و بهتر ناباروری با دلایل گوناگون اقدام نمود.

در پژوهش حاضر تلاش شد تا با استفاده از الگوریتم خوشه‌بندی از مجموعه الگوریتم‌های غیرنظارتی، داده‌های ناباروری را تحلیل کرد. پژوهش‌های زیادی با استفاده از الگوریتم‌های داده‌کاوی در حوزه ناباروری انجام شده است که هر کدام بر فاکتور خاصی تمرکز دارد.

در پژوهشی با هدف بررسی میزان موفقیت درمان با روش IUI، عوامل مختلفی مانند طول مدت نازایی، مشخصات زوجین، وضعیت آزمایش اسپرم و غیره روی داده‌ها، به این نتیجه رسیده‌اند که بهترین نتیجه درمانی برای نازایی با علت تخمدانی، با روش IUI به دست می‌آید[5].

در پژوهشی دیگر برای تعیین میزان موفقیت درمان با روشIVF، عوامل مختلف موثر در باروری بررسی شده‌اند و نتایج نشان داد که پاسخ تخمدان به تحریک تخمک‌گذاری و تعداد جنین منتقل شده، از عوامل مهم و موثر در پیش‌بینی نتایج IVF است[6].

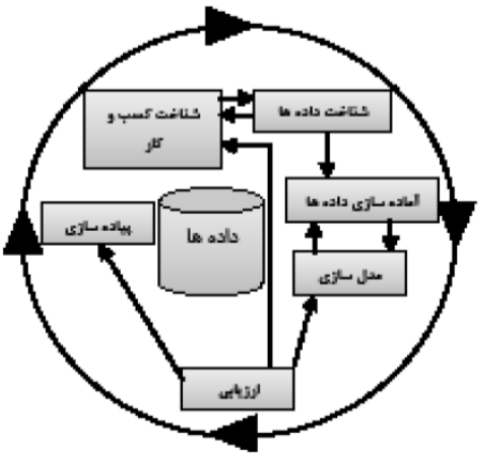
در پژوهشی با بررسی تعداد جنین منتقل‌شده در سیکل‌های ART، مشخص شد که با انتقال ۲ یا ۳ جنین به جای ۱ جنین، احتمال بارداری به میزان معنی‌داری افزایش می‌یابد از طرفی بیشتر از آن، میزان افزایش احتمال چشمگیر نیست و به ریسک چندقلوزایی نمی‌ارزد[7]. لازم به ذکر است که پژوهش‌های بیشتری در این زمینه وجود دارد که تنها به میزان کوچکی از آنها اشاره شده است.

در همین راستا، علم داده‌کاوی با استفاده از مفاهیم جدید هوش مصنوعی و علوم آماری به کمک علوم مختلف از جمله علم ناباروری آمده است تا به تفسیر نتایج و تحلیل‌های مناسب داده‌ها کمک کند و بتواند الگوها و دانش‌های نهفته در داده‌ها را استخراج کند تا با امکان‌پذیری پیش‌بینی نتایج، بتوان از آن در تصمیم‌گیری بالینی برای بیماران نابارور بهره گرفت.

این مطالعه با هدف تحلیل عوامل موثر بر نتایج روش انتقال داخل‌رحمی اسپرم با استفاده از خوشه‌بندی انجام شد.

ابزار و روش‌ها

این مطالعه براساس مدل شناخته‌شده فرآیند استاندارد صنعت[8] (CRISP) انجام شد، این مدل در پژوهش‌های داده‌کاوی استفاده می‌شود (شکل ۱).



شکل ۱) روش‌شناسی CRISP

فرآیندهایی که براساس این روش، در این پژوهش دنبال شد، شامل موارد زیر بودند:

شناخت کسب و کار: در این گام از تحلیل، فرآیندهای موجود در بیمارستان شناخته و بررسی شدند. این مرحله از مهم‌ترین مراحل پژوهش است و در طول اجرای پروژه بارها به این مرحله باز گشته شد، تا شناخت از سیستم بیمارستان کامل‌تر و نتایج حاصل از داده‌کاوی را با آن ارزیابی شد.

شناخت داده‌ها: بیماران برای درمان‌های مختلف از جمله ICSI، IUI، GIFT، ZIFT و IVF به بیمارستان صارم مراجعه می‌کنند. اطلاعات شخصی زوجین و نتایج آزمایشات در پرونده‌های آنها ثبت و نگهداری می‌شود. در این فاز، با مشورت و کمک تیم بیمارستان کلیه داده‌های موجود در بیمارستان صارم شناخته شد. این شناخت کلیه رکوردهای داده‌ای، صفات داده‌ها و همچنین آشنایی با مفاهیم انواع روش‌های درمانی (شامل انواع آزمایشات)، را دربرمی‌گیرد.

آماده‌سازی داده‌ها: از فازهای مهم این پژوهش، مرحله پاک‌سازی و آماده‌سازی داده‌ها برای انجام فرآیند داده‌کاوی بود که به‌عنوان مرحله پیش‌پردازش داده‌ها شناخته شد. در این مرحله، برای سهولت در انجام پژوهش و هدفمندبودن نتایج، از بین رکوردهای ثبت‌شده در بیمارستان که در زمینه رشته‌های مختلف درمانی و زمان‌های متفاوت بود، داده‌های مورد نیاز انتخاب شد و تحلیل و بررسی روی این دسته از داده‌ها آغاز شد. در این فاز، برای محدودکردن تحلیل، یکی از انواع تکنیک‌های ناباروری مدنظر قرار گرفت تا داده‌های آن تحلیل شود. داده‌های مربوط به بیمارانی که عمل IUI روی آنها انجام شده بود، بررسی و الگوریتم خوشه‌بندی روی آنها پیاده و تحلیل‌های لازم انجام شد. همچنین از بین فیلدهای موجود در پایگاه داده، برخی پارامترهای موثرتر که با آزمایش‌های مکرر پژوهش‌های گذشته در این زمینه و مشورت با افراد مجرب این رشته به دست آمده بود، انتخاب و به‌کار برده شد (جدول ۱).

پس از مرحله شناخت سیستم بیمارستان و آشنایی با فرآیندهایی که در طول درمان برای بیماران انجام می‌شود و در مراحل بعد، شناخت داده‌ها و سپس انتخاب داده‌های مورد نیاز از حجم عظیم داده‌ها به مرحله پاک‌سازی و پیش‌پردازش داده‌ها انجام شد. در این مرحله، داده‌ها بررسی، طبقه‌بندی، همین‌طور حذف یا اصلاح شدند (اصلاح داده‌های اشتباه یا مواردی که به‌ازای برخی ویژگی‌ها مقدار مشخصی نداشتند). نتیجه حاصل از انجام پاک‌سازی داده‌ها، ۴۰۰ پرونده بین سال ۱۳۷۶ تا سال ۱۳۸۸ از بیماران بیمارستان صارم بود.

جدول ۱) پارامترهای انتخاب‌شده

پارامترها
سن
شاخص توده بدنی (BMI) (kg/m2)
طول ناباروری
ناباروری اولیه
ناباروری ثانویه
عامل مردانه
عامل تخمک‌گذاری
ضخامت آندومتر
تعداد اسپرم‌ها بعد از شست‌وشو
تعداد اسپرم‌های نرمال بعد از شست‌وشو
تعداد اسپرم‌های Full بعد از شست‌وشو
تعداد اسپرم‌های sluggish بعد از شست‌وشو
وضعیت نرمال یا غیرنرمال‌بودن اسپرم
تعداد فولیکول

در مرحله بعدی، مدل خوشه‌بندی انتخاب شده روی نمونه آماری آماده‌شده، اجرا شد، تا در نتیجه آن دانش و الگوهای نهفته در این داده‌ها استخراج شود.

مدل‌سازی: در این مرحله، روش مدل‌سازی داده‌های موجود انتخاب شد. برای مدل‌سازی داده‌ها، یکی از الگوریتم‌های غیرنظارتی (توصیفی) استفاده شد. از رایج‌ترین الگوریتم‌های توصیفی، الگوریتم خوشه‌بندی است که در این پژوهش نیز برای تحلیل داده‌های ناباروری به‌کار رفت. الگوریتم‌های توصیفی، به‌منظور کشف الگوهای مشابه بین دسته‌های مختلف داده‌ها استفاده می‌شود. مساله اساسی در الگوریتم خوشه‌بندی به‌کار رفته در این پژوهش، توزیع داده‌ها به K گروه مختلف که داده‌های هر گروه با هم مشابه و داده‌های گروه‌های مختلف با هم نامتشابه باشند که به آن "K-means" می‌گویند[9]. با استفاده از این الگوریتم غیرنظارتی، می‌توان به یک تحلیلی کلی از داده‌ها و نوعی تقسیم‌بندی روی داده‌ها رسید. در استفاده از الگوریتم خوشه‌بندی، نیازمند انتخاب یک شاخص برای انجام آن هستیم. در این پژوهش، از شاخص دیویس- بولدین استفاده شد. با کاهش مقدار این شاخص، تعداد خوشه شبیه به آن شاخص، نتیجه بهتری ارایه می‌کند.

برای رسیدن به تحلیل مناسب و تفسیر بهتر نتایج، چندین‌بار الگوریتم خوشه‌بندی را با شاخص‌های مختلف روی داده‌ها اجرا شد. این روش، علاوه بر راهنمایی برای انتخاب بهترین شاخص خوشه‌بندی، کمک زیادی به تصحیح ویژگی‌ها و داده‌ها کرد. الگوریتم‌های خوشه‌بندی، چندین‌بار و با استفاده از شاخص‌های مختلف روی داده‌ها اجرا و در نهایت مشخص شد که شاخص دان و دیویس، شاخص‌های بهتری در اجرای الگوریتم‌ها روی این داده‌ها بود.

شاخص دان (Dunn Index): این شاخص یکی از رایج‌ترین شاخص‌ها در ارزیابی تعداد خوشه‌ها است که به‌صورت زیر تعریف می‌شود:

معادله ۱.

$$V(\delta) = \frac{min_h\, k = 1, \dots K\; d_c(C_k, C_h)}{\max k = 1, \dots k\Delta(C_k)}$$

در معادله ۱، عبارت $d_c(C_k,C_h)$ فاصله بین دو خوشه k و h را محاسبه می‌کند که در صورت فرمول قرار دارد و عبارت $\Delta(C_k)$ به محاسبه چگالی خوشه k می‌پردازد، به این معنی که نزدیکی نمونه‌های داخل خوشه k را به‌دست می‌آورد. با توجه به روش‌های مختلفی که برای محاسبه فاصله بین دو خوشه و اندازه خوشه وجود دارد، می‌توان به شاخص‌های مختلفی از خانواده دان رسید.

شاخص دیویس- بولدین (Davies-Bouldin Index): این شاخص تابعی از نسبت جمع پراکندگی داخل خوشه به جدایی خوشه‌ها از هم است[10]. فرض کنید در معادله ۲، c_i نشان‌دهنده خوشه و X_j بُعد ویژگی‌هایی باشد که به خوشه نسبت داده شده است:

معادله ۲

$$S_i = \sqrt[q]{\frac{1}{T_i}\sum_{j=1}^{T_i}|X_j - A_i|^q}$$

در معادله بالا A_i مرکز خوشه c_i و T_i نیز سایز خوشه است که در نتیجه S_i میزان پراکندگی در خوشه را نشان می‌دهد و اغلب به دلیل اینکه از فاصله اقلیدسی استفاده می‌شود، q نیز مقدار ۲ را دارد، اما همان‌طور که در شاخص دان نیز می‌توانست از فرمول‌های دیگری غیر از اقلیدسی استفاده کرد در این فرمول نیز می‌توان استفاده کرد:

معادله ۳.

$$S_i = \sqrt[q]{\frac{1}{T_i}\sum_{j=1}^{T_i}|X_j - A_i|^q}$$

در معادله ۳ نیز $M_{i,j}$ برای محاسبه میزان فاصله بین خوشه‌های C_i و C_j است و $a_{k,j}$ نیز نشان‌دهنده kاُمین عنصر در A_i است.

معادله ۴.

$$S_i = \sqrt[q]{\frac{1}{T_i}\sum_{j=1}^{T_i}|X_j - A_i|^q}$$

در معادله ۴ نیز $R_{i,j}$ میزانی است که مشخص می‌کند خوشه چقدر مناسب است. در نهایت شاخص دیویس- بولدین به‌صورت زیر تعریف می‌شود:

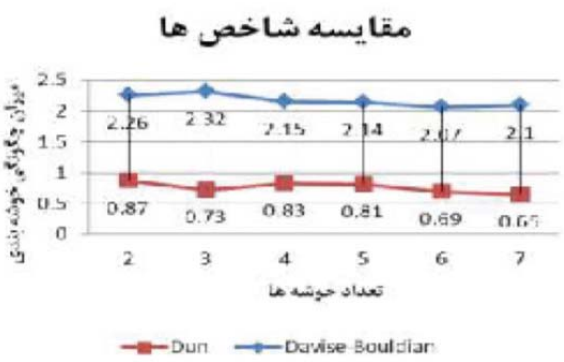
معادله ۵.

$$S_i = \sqrt[q]{\frac{1}{T_i}\sum_{j=1}^{T_i}|X_j - A_i|^q}$$

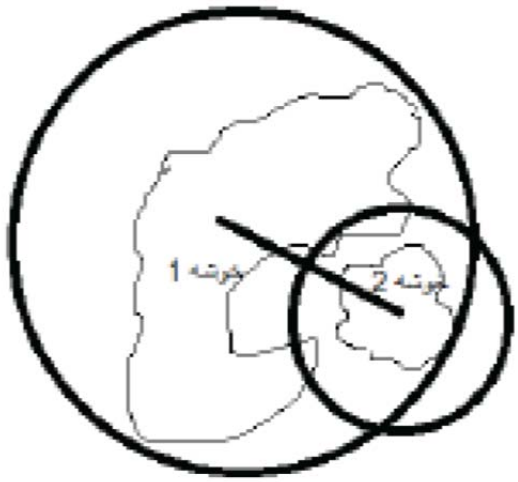
با توجه به معادله ۵ صورت فرمول مربوطه، نشان‌دهنده چگال‌بودن خوشه‌ها و مخرج فرمول نشان‌دهنده فاصله خوشه‌ها از هم است، در نتیجه صورت هرچه کوچک‌تر و مخرج هرچه بزرگ‌تر باشد، مطلوب‌تر خواهد بود. پس مقدار DB هرچه کمتر باشد، مطلوب‌تر است.

انتخاب تعداد خوشه بهینه: با آنچه در مورد فرمول‌های این دو شاخص در بالا ذکر شد، این معادلات در نرم‌افزار مربوطه پیاده و خوشه‌بندی تا ۷ خوشه روی داده‌های IUI انجام شد که نتایج زیر را در برداشت.

مطابق آنچه در مورد الگوریتم‌های دان و دیویس گفته شد چنان چه از مرکز خوشه‌ها، دایره‌هایی را رسم کنیم، این طور به‌نظر می‌آید که خوشه‌ها با هم تلاقی دارند (نمودار۱ و شکل ۲).



نمودار۱) مقایسه شاخص دان و دیویس



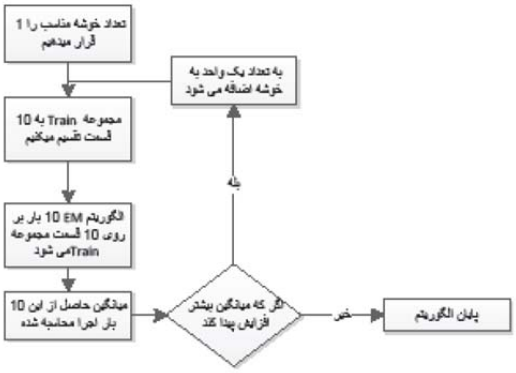
شکل ۲) تلاقی خوشه‌ها

از این رو شاخص دان نتیجه مناسبی را برای این خوشه‌بندی ارایه نکرد، زیرا همان‌طور که گفته شد، این شاخص، مبتنی بر فاصله بین خوشه و قطر خوشه است که این مقدار هر چه بیشتر باشد، بهتر است. در مقایسه، شاخص دیویس، بر نسبت چگالی به فاصله متمرکز است و تنها به فاصله اکتفا نمی‌کند، در نتیجه حساسیت کمتری به تلاقی بین خوشه‌ها دارد. به‌نظر می‌رسد که خوشه‌ها نسبت به همدیگر تلاقی دارند، درصورتی‌که این طور نیست (شکل ۲).

با درنظر گرفتن تعداد بهینه‌ای که هریک از این شاخص‌ها نشان دادند، شاخص دان، ۲ خوشه را بهترین و شاخص دیویس، ۶ خوشه را به‌عنوان شاخص‌ترین تعداد خوشه در نظر گرفت و توجه به بحث، ۶ خوشه به‌عنوان تعداد خوشه بهینه در نظر گرفته شد. از آنجایی که در الگوریتم K-means ممکن است در بهینه محلی قرار بگیریم بر آن شدیم تا از الگوریتم دیگری به نام EM (Expectation Maximization) استفاده کنیم[11] تا از نتیجه حاصل از الگوریتم K-means و شاخص DB مطمئن شویم. نتایج حاصل از اجرای

الگوریتم EM نیز ۶ خوشه را به ما ارایه کرد. در این الگوریتم به ازای هر نمونه یک توزیع احتمال اختصاص می‌دهند که احتمال هر خوشه را مشخص می‌کند. این الگوریتم می‌تواند تصمیم درباره تعداد خوشه‌های مناسب را با استفاده از الگوریتم داخلی[11] اعتبار سنجی متقابل (Cross validation) که در آن پیاده‌سازی شده، انجام دهد. برای اجرای الگوریتم، در برنامه WEKA پیاده‌سازی و اجرا شد. همان‌طور که بیان شد، الگوریتم Cross validation تعداد خوشه‌های مناسب را تشخیص می‌دهد که این کار به صورت زیر انجام می‌شود.

این دو الگوریتم با دو رویکرد متفاوت نتیجه یکسان را ارایه کردند اما در مرحله آخر از اطمینان قطعی از تعداد خوشه‌های مناسب نتایج این دو الگوریتم با افراد مجرب این حوزه در بیمارستان صارم بررسی شد تا از صحت اجرای این الگوریتم‌ها اطمینان حاصل شود (شکل ۳).



شکل ۳) روند محاسبه تعداد خوشه مناسب در الگوریتم EM

یافته‌ها

پس از ۷ بار اجرای الگوریتم K-means که برای خوشه‌بندی روی داده‌ها انتخاب شده بود و با توجه به انتخاب شاخص دیویس- بولدین برای انتخاب تعداد خوشه‌های بهینه، تعداد ۶ خوشه انتخاب و پراکندگی بیماران در هر خوشه مشخص شد (جدول ۲).

جدول ۲) پراکندگی بیماران در هر خوشه

اسم خوشه	فراوانی در K-means	فراوانی در EM	درصد فراوانی	درصد فراوانی خوشه‌بندی الگوریتم EM
خوشه ۱	۱۲۸	۱۵۷	۳۲	۳۹
خوشه ۲	۳۲	۳۰	۸	۸
خوشه ۳	۵۲	۴۷	۱۳	۱۲
خوشه ۴	۲۷	۱۳	۷	۳
خوشه ۵	۹۱	۸۱	۲۳	۲۰
خوشه ۶	۷۰	۷۲	۱۷	۱۸

در ادامه هرکدام از ۶ خوشه انتخاب‌شده، تحلیل شده‌اند.

نتایج به‌دست‌آمده از خوشه یک را در جدول ۳ مشاهده می‌کنید، تعداد ۶ نمونه از نمونه‌هایی که در خوشه بالا قرار داشتند نتیجه‌ای در پایگاه داده برای موفقیت و عدم موفقیت آنها ثبت نشده بود (جدول ۴).

نتایج به‌دست‌آمده از خوشه دوم نیز در جدول ۵ و نتایج موفقیت و عدم موفقیت در این خوشه نیز به شرح جدول ۶ است.

جدول ۳) نتایج حاصل از خوشه اول

پارامتر	ارزش پارامتر در خوشه
عامل تخمک‌گذاری	%۱۰۰
عامل مردانه	%۹۳/۰
سن	(۱۹ تا ۴۹) میانگین ۲۷سال
BMI<۱۸/۵	%۳/۰
۱۸/۵<BMI<۲۴/۹	%۴۸/۰
۲۵<BMI<۲۹/۹	%۳۲/۰
BMI>۳۰	%۱۷/۰
طول ناباروری	(۱ تا ۱۶) میانگین ۵/۴سال
ناباروری اولیه	%۱۰۰
ناباروری ثانویه	۰
ضخامت آندومتر	(۳/۷ تا ۱۴) ، میانگین ۸میلی‌متر
تعداد اسپرم بعد از شست‌وشو	(۰/۲۵ تا ۱۶) میانگین ۷میلیون
تعداد اسپرم‌های نرمال بعد از شست‌وشو	(۰/۱۲۵ تا ۱۰/۵۶) میانگین ۴/۵
تعداد اسپرم‌های FULL بعد از شست‌وشو	(۰ تا ۱۲) میانگین ۳/۹میلیون
تعداد اسپرم‌های کند بعد از شست‌وشو	(۰ تا ۷/۰۸) میانگین ۳/۰۲۵میلیون
Terato Spermia	%۸۴/۴
Asterato Spermia	%۳/۱
Pyo spermia	%۱۱/۷
Spermogram Normal	%۰/۸

جدول ۴) نتایج خوشه اول (%)

پارامتر	ارزش پارامتر در خوشه
عدم موفقیت	۸۰/۱
موفقیت	۱۵/۳
نامشخص	۴/۶

جدول ۵) نتایج حاصل از خوشه دوم

پارامتر	ارزش پارامتر در خوشه
عامل تخمک‌گذاری	۰
عامل مردانه	%۹۳/۸
سن	(۲۲ تا ۴۴) میانگین ۳۲سال
BMI<۱۸/۵	۰
۱۸/۵<BMI<۲۴/۹	%۷۱/۹
۲۵<BMI<۲۹/۹	۰
BMI>۳۰	%۲۸/۱
طول ناباروری	(۰/۵ تا ۱۲) میانگین ۲/۷سال
ناباروری اولیه	۰
ناباروری ثانویه	%۱۰۰
ضخامت آندومتر	(۵ تا ۱۱) میانگین ۷میلی‌متر
تعداد اسپرم بعد از شست‌وشو	(۲ تا ۱۴) میانگین ۷/۲میلیون
تعداد اسپرم‌های نرمال بعد از شست‌وشو	(۱ تا ۹/۹) میانگین ۴/۶میلیون
تعداد اسپرم‌های FULL بعد از شست‌وشو	(۰ تا ۱۰/۶) میانگین ۴/۲میلیون
تعداد اسپرم‌های کند بعد از شست‌وشو	(۰/۵ تا ۵/۰۱) میانگین ۳/۰۳میلیون
Terato Spermia	%۱۰۰

جدول ۶) نتایج خوشه دوم (%)

پارامتر	ارزش پارامتر در خوشه
عدم موفقیت	۹۰/۷
موفقیت	۹/۳
نامشخص	۰

نتایج به‌دست‌آمده از خوشه سوم نیز به شرح جدول ۷ است و نتایج موفقیت و عدم موفقیت در جدول ۸ آورده شده است.

جدول ۷) نتایج حاصل از خوشه سوم

پارامتر	ارزش پارامتر در خوشه
عامل تخمک‌گذاری	%۵/۸
عامل مردانه	%۱۰۰
سن	(۲۰ تا ۴۳) میانگین ۲۸سال
BMI<۱۸/۵	۰
۱۸/۵<BMI<۲۴/۹	%۴۲/۳
۲۵<BMI<۲۹/۹	%۵۰
BMI>۳۰	%۷/۷
طول ناباروری	(۱ تا ۱۲) میانگین ۳/۴سال
ناباروری اولیه	%۶۵/۴
ناباروری ثانویه	%۳۴/۶
ضخامت آندومتر	(۲/۵ تا ۱۳) میانگین ۷/۹میلی‌متر
تعداد اسپرم بعد از شست‌وشو	(۱/۲۵ تا ۱۴/۵) میانگین ۴/۸میلیون
تعداد اسپرم‌های نرمال بعد از شست‌وشو	(۰/۴۹ تا ۹/۷) میانگین ۲/۹میلیون
تعداد اسپرم‌های FULL بعد از شست‌وشو	(۰/۹ تا ۴) میانگین ۲/۱میلیون
تعداد اسپرم‌های کند بعد از شست‌وشو	(۰ تا ۶) میانگین ۲/۵میلیون
Terato Spermia	%۱۰۰

جدول ۸) نتایج خوشه سوم

پارامتر	ارزش پارامتر در خوشه (%)
عدم موفقیت	۸۸/۵
موفقیت	۱۰/۸
نامشخص	۰/۷

براساس جدول ۸ تعداد ۱ نمونه از نمونه‌هایی که در خوشه بالا قرار داشتند نتیجه‌ای در پایگاه داده برای موفقیت و عدم موفقیت آنها ثبت نشده بود.

نتایج به‌دست‌آمده از خوشه چهارم در جدول ۹ و نتایج موفقیت و عدم موفقیت در این خوشه نیز در جدول ۱۰ ذکر شده است.

جدول ۹) نتایج حاصل از خوشه چهارم

پارامتر	ارزش پارامتر در خوشه
عامل تخمک‌گذاری	%۸۸/۸
عامل مردانه	%۴۰/۷
سن	(۱۹ تا ۴۲) میانگین ۲۹سال
BMI<۱۸/۵	۰
۱۸/۵<BMI<۲۴/۹	%۱۴/۸
۲۵<BMI<۲۹/۹	%۷/۴
BMI>۳۰	%۷۷/۷
طول ناباروری	(۱ تا ۱۶) میانگین ۵/۱سال
ناباروری اولیه	%۷۰/۴
ناباروری ثانویه	%۲۹/۶
ضخامت آندومتر	(۱۱تا ۴/۵) میانگین ۸میلی‌متر
تعداد اسپرم بعد از شست‌وشو	(۲/۵ تا ۱۲/۵) میانگین ۷/۵میلیون
تعداد اسپرم‌های نرمال بعد از شست‌وشو	(۱/۷۵ تا ۹/۷۵)میانگین ۴/۶میلیون
تعداد اسپرم‌های FULL بعد از شست‌وشو	(۰ تا ۱۰/۶۲) میانگین ۴/۵میلیون
تعداد اسپرم‌های کند بعد از شست‌وشو	(۱ تا ۵/۴) میانگین ۲/۷میلیون
Terato Spermia	%۷/۴
Pyo spermia	%۳۷/۰
Spermogram Normal	%۵۵/۵

جدول ۱۰) نتایج خوشه چهارم (%)

پارامتر	ارزش پارامتر در خوشه
عدم موفقیت	۸۸/۸
موفقیت	۱۱/۱
نامشخص	۰

نتایج به‌دست‌آمده از خوشه پنجم نیز به شرح جدول ۱۱ است؛ همچنین نتایج موفقیت و عدم موفقیت در این خوشه نیز به شرح جدول ۱۲ است.

جدول ۱۱) نتایج حاصل از خوشه پنجم

پارامتر	ارزش پارامتر در خوشه
عامل تخمک‌گذاری	۰
عامل مردانه	%۹۶/۷
سن	(۲۰ تا ۴۰) میانگین ۲۸سال
BMI<۱۸/۵	۰
۱۸/۵<BMI<۲۴/۹	%۶۰/۴
۲۵<BMI<۲۹/۹	%۲۸/۵
BMI>۳۰	%۱۰/۹
طول ناباروری	(۰/۵تا ۱۲) میانگین ۳/۷سال
ناباروری اولیه	%۱۰۰
ناباروری ثانویه	۰
ضخامت آندومتر	(۵ تا ۱۶) میانگین ۷/۸میلی‌متر
تعداد اسپرم بعد از شست‌وشو	(۰/۷ تا ۱۳/۵) میانگین ۶/۵میلیون
تعداد اسپرم‌های نرمال بعد از شست‌وشو	(۰/۳ تا ۹) میانگین ۴/۱میلیون
تعداد اسپرم‌های FULL بعداز شست‌وشو	(۰ تا ۱۰/۸) میانگین ۳/۴میلیون
تعداد اسپرم‌های کند بعد از شست‌وشو	(۰/۲ تا ۵/۵) میانگین ۳میلیون
Astheno Spermia	%۲/۲
Terato Spermia	%۹۶/۷
Asterato Spermia	%۱/۱

جدول ۱۲) نتایج خوشه پنجم

پارامتر	ارزش پارامتر در خوشه (%)
عدم موفقیت	۸۹/۰
موفقیت	۷/۱
نامشخص	۳/۹

جدول ۱۳) نتایج حاصله از خوشه ششم

پارامتر	ارزش پارامتر در خوشه
عامل تخمک‌گذاری	%۵۱/۴
عامل مردانه	%۹۲/۹
سن	(۲۱ تا ۴۱) میانگین ۳۱سال
BMI<۱۸/۵	۰
۱۸/۵<BMI<۲۴/۹	%۱۱/۴
۲۵<BMI<۲۹/۹	%۷۴/۳
BMI>۳۰	%۱۴/۳
طول ناباروری	(/۵ ۰تا ۱۱) میانگین ۳سال
ناباروری اولیه	۰
ناباروری ثانویه	%۱۰۰
ضخامت آندومتر	(۳ تا ۱۴) میانگین ۷/۵میلی‌متر
تعداد اسپرم بعد از شست‌وشو	(۱/۲۵ تا ۱۵) میانگین ۷میلیون
تعداد اسپرم‌های نرمال بعد از شست‌وشو	(۰/۵ تا ۹/۹) میانگین ۴/۳میلیون
تعداد اسپرم‌های FULL بعد از شست‌وشو	(۰ تا ۱۰/۲) میانگین ۳/۶میلیون
تعداد اسپرم‌های کند بعد از شست‌وشو	(۰/۵ تا ۴/۸) میانگین ۲/۹میلیون
Astheno Spermia	%۸/۵
Terato Spermia	%۸۵/۷
Oligo asterato.s	%۱/۴
Asterato Spermia	%۱/۴
Spermogram Normal	%۲/۹

نتایج به‌دست‌آمده از خوشه ششم نیز به شرح جدول ۱۳ است. براساس جدول۱۳، ۵ نمونه از نمونه‌هایی که در خوشه بالا قرار داشتند نتیجه‌ای در پایگاه داده برای موفقیت و عدم موفقیت آنها

ثبت نشده بود. نتایج موفقیت و عدم موفقیت در این خوشه نیز در جدول ۱۲ است.

جدول ۱۴) نتایج خوشه ششم (%)

پارامتر	ارزش پارامتر در خوشه
عدم موفقیت	۸۰/۰
موفقیت	۱۵/۳
نامشخص	۴/۷

براساس جدول ۱۴ تعداد ۶ نمونه از نمونه‌هایی که در خوشه بالا قرار داشتند نتیجه‌ای در پایگاه داده برای موفقیت و عدم موفقیت آنها ثبت نشده بود.

بحث

پژوهش‌های زیادی با استفاده از الگوریتم‌های داده‌کاوی در حوزه ناباروری انجام شده است که در هریک از آنها، بر فاکتور خاصی تمرکز شده است. نمونه‌هایی از فعالیت‌های انجام‌شده در این حوزه در ادامه ذکر خواهد شد. بررسی میزان موفقیت درمان با روش IUI در این پژوهش، با بررسی عوامل مختلفی همچون طول مدت نازایی، مشخصات زوجین، وضعیت آزمایش اسپرم و غیره روی داده‌ها، به این نتیجه رسیدند که بهترین نتیجه درمانی برای نازایی با علت تخمدانی، با روش IUI به دست می‌آید[5]. در این پژوهش با بررسی غیرنظارتی بر عوامل مختلف در درمان IUI به‌دسته‌بندی‌کردن بیمارانی که عمل IUI روی آنها انجام شده است پرداخته و نیز میزان موفقیت و عدم موفقیت هر دسته را مورد بررسی قرار گرفت. در پژوهش دیگری مشابه با پژوهش حاضر اما روی روش IVF، میزان موفقیت درمان با روش IVF و عوامل مختلف موثر در باروری بررسی شد، یافته‌ها نشان داد که پاسخ تخمدان به تحریک تخمک‌گذاری و تعداد جنین منتقل‌شده از عوامل مهم و موثر در پیش‌بینی نتایج IVF است[5]. در سال ۲۰۰۸ پژوهش دیگری مرتبط با میزان احتمال حاملگی با تعداد جنین منتقل‌شده انجام شد که با بررسی تعداد جنین منتقل‌شده در سیکل‌های ART مشخص شد با انتقال ۲ یا ۳ جنین به جای ۱ جنین، احتمال حاملگی به میزان معنی‌داری افزایش می‌یابد و بیشتر از آن، میزان افزایش احتمال، چشمگیر نیست و به ریسک چند قلوزایی نمی‌ارزد. در این پژوهش با استفاده از تکنیک‌های شبکه‌های بیزین احتمال چند قلوزایی بررسی شد[7].

در حوزه‌های دیگر مانند پیش‌بینی در زمینه درمان IVF نیز پژوهش‌هایی صورت گرفته است مانند پژوهشی در سال ۲۰۱۰ که براساس نمونه‌برداری، بررسی ویژگی‌ها و پاک‌سازی داده‌ها، ۳۹ ویژگی یافت شد، به‌منظور افزایش دقت، ترکیب درخت تصمیم و الگوریتم ژنتیک را استفاده کرده‌اند[12]. از سیستم‌های دیگر که در حوزه ناباروری می‌توان به آن اشاره کرد سیستم پیش‌بینی نوع مناسب درمان IVF با استفاده از CBR است، این سیستم TA3IV نام دارد. هدف پژوهش این است که توسط CBR از تجربیات ذخیره شده یک راه ممکن برای IVF به‌منظور افزایش فاکتورهای موفقیت ارایه دهد. پس از آنکه دانش، سیستم مبتنی بر جمعیت مناسبی از نمونه‌ها شد، قابلیت استفاده استخراج دانش‌های مورد استفاده را از انبوهی از داده‌ها پیدا می‌کند. در این پژوهش سیستم TA3IV که یک سیستم CBR براساس دانش‌کاوی است، توضیح داده شده است. این سیستم می‌تواند در کنار پزشکان یک پیشنهاددهنده باشد[13]. در پژوهشی، ۲۴۵۰ زوج بررسی شدند، که مبنایی مناسب برای انتخاب ویژگی در این پژوهش نیز بود، اما فقط به بررسی آماری بیماران اشاره شد. از جمله ویژگی‌های انتخاب شده می‌توان به BMI، نوع ناباروری، طول ناباروری، سن و غیره نام برد[14]. تفاوت عمده در این پژوهش این است که با وجود فیلد موفقیت و عدم موفقیت در انجام عمل IUI ابتدا آنها را از ویژگی‌ها حذف و به بررسی وضعیت بیماران پرداخته شد، زیرا اهم بررسی در به‌دست‌آوردن شناخت کلی از بیماران بیمارستان بوده و در ادامه بعد از تحلیل داده‌ها بتوان براساس خوشه‌بندی هرکدام از بیماران به میزان موفقیت و عدم موفقیت آنها پرداخته شود و از این بستر بتوان در آینده برای ایجاد سیستم تصمیم‌یار استفاده کرد. افزایش روزافزون پیشرفت در زمینه تکنولوژی اطلاعات، موجب ایجاد رشد چشم‌گیری در علوم شده است. یکی از این رشته‌ها، علم پزشکی است. استفاده از تکنیک‌های داده‌کاوی در این شاخه از علم، موجب شده که در کلیه مباحث، به‌خصوص بحث ناباروری، بتوان سیستم‌های پزشک‌یار طراحی کرد. عوامل مختلفی موجب بروز ناباروری می‌شود که می‌توان آنها را به‌طور کلی به دو دسته عوامل مردانه و عوامل زنانه تقسیم کرد. هرکدام از این عوامل فاکتورهای زیادی را شامل می‌شوند و درمان خاص خود را دارند. روش‌های درمانی از جمله IVF، IUI و غیره در این حوزه استفاده می‌شود. در این پژوهش با خارج‌کردن ویژگی موفقیت و عدم موفقیت سعی شد با استفاده از تحلیل‌های داده‌کاوی روش‌های غیرنظارتی، داده‌های این حوزه تحلیل شود تا بتوان به پزشکان در تصمیم‌گیری برای درمان بیماران کمک کرد. شانس موفقیت در روش IUI ۱۰ تا ۱۵% اعلام شده است که در پژوهش حاضر، براساس ۴۰۰ رکورد مناسب و پاک‌سازی‌شده برای اجرا، الگوریتم‌ها را به ۶ خوشه تقسیم کرده، عواملی که با توجه به بررسی‌ها انتخاب و در هرکدام از خوشه‌ها میزان موفقیت و عدم موفقیت نیز بررسی شد. این نتایج می‌تواند به‌عنوان منبعی برای بررسی بیماران جدید باشد تا بتوان میزان موفقیت و عدم موفقیت روش IUI را روی آنها دقیق‌تر و علمی‌تر بررسی کرد. در این پژوهش سعی شد، در ابتدا بدون درنظرگرفتن موفقیت یا عدم موفقیت هر بیمار به یک توصیف کلی از بیماران براساس ویژگی‌های انتخاب‌شده پرداخته شود و جدای از تقسیم‌بندی بیماران به دسته‌های موفق و ناموفق بتوان به تفسیر درستی از داده‌ها رسید. میتوان در آینده، سیستم تصمیم‌یاری طراحی شود تا بتوان با تفسیر درستی که از داده‌ها انجام شده به ارایه راه حل بپردازد و به یاری پزشکان به‌منظور انتخاب درست روش درمان کمک کند.

نتیجه‌گیری

عواملی مانند سن، حجم توده بدنی، نوع ناباروری، علت ناباروری و غیره می‌توانند میزان موفقیت روش IUI را تعیین کنند.

تشکر و قدردانی: از زحمات جناب آقای دکتر صارمی ریاست محترم بیمارستان و جناب آقای دکتر صالحیان و همچنین کارکنان متخصص و با کفایت بیمارستان فوق تخصصی زنان صارم که در مراحل مختلف این پژوهش ما را یاری کردند، به‌ویژه سرکار خانم غفاری و سرکار خانم شامی کمال تشکر و قدردانی را داریم.

تاییدیه اخلاقی: موردی از سوی نویسندگان ذکرنشده است.

تعارض منافع: موردی وجود نداشته است.

منابع مالی: توسط مرکز تحقیقات باروری و ناباروری صارم و پژوهشکده سلولی- مولکولی و سلول‌های بنیادی صارم تامین شده است.

سهم نویسندگان: سمیه علیزاده (نویسنده اول)، نگارنده مقدمه/روش‌شناس/پژوهشگر اصلی/نگارنده بحث (۴۰%)؛ محسن اصغری (نویسنده دوم)، نگارنده مقدمه/پژوهشگر اصلی/ تحلیل‌گر آماری/ نگارنده بحث (۳۰%)؛ محمدکاظم حسینی (نویسنده سوم)، نگارنده مقدمه/روش‌شناس/پژوهشگر اصلی/تحلیل‌گر آماری/نگارنده بحث (۳۰%)

منابع

1- Chakrabarti S, Ester M, Fayyad U, Gehrke J, Han J, Morishita S, et al. Data Mining Curriculum: A Proposal (Version 1.0) [Internet]. London: The community for data mining, data science and analytics, SIGKDD; 1999. [updated 2006 Apr 30; cited 2007 Dec 14]. Avilable from: www.kdd.org/exploration_files/CURMay06.pdf.

2- Berry GT, Baker L, Kaplan FS, Witzleben CL. Diabetes-like renal glomerular disease in Fanconi-Bickel syndrome. Pediatr Nephrol. 1995;9(3):287-91.

3- Han J, Kamber M. Data Mining: Concepts and Techniques. 2th ed. Burlington: Morgan Kaufmann; 2011.

4- Saremi AT. Infertility Guideline. Tehran: Sarem Research Center; 2009. [Persian]

5- Vahid Roudsari F, Ayati S, Mirzaeeyan S, Shakeri MT, Akhtardel H. Fertility outcome after IVF and related factors. J Gorgan Univ Med Sci. 2009;11(3):42-6. [Persian]

6- Morales DA, Bengoetxea E, Larranaga P, Garcia M, Franco Y, Fresnada M, et al. Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. Comput Methods Programs Biomed. 2008;90(2):104-16.

7- Sohrabvand F, Shariat M, Fotoohi Ghiam N, Hashemi M. The relationship between number of transferred embryos and pregnancy rate in ART cycles. Tehran Univ Med J. 2009;67(2):132-6. [Persian]

8- Gazanfari M, Alizadeh S, Teimourpur B. Data Mining and Knowledge Discovery. Tehran: Iran University of Science and Industry Press; 2011.

9- Han L, Zhong Y, Huang B, Han L, Pan L, Xu X, et al. Sodium butyrate activates erythroid-specific 5-aminolevulinate synthase gene through Sp1 elements at its promoter. Blood Cells Mol Dis. 2008;41(2):148-53.

10- Gonzalez T, Marggie D. A comparison in cluster validation techniques [Dissertation]. Puerto Rico: University of Puerto Rico; 2006.

11- Witten IH, Frank E, Hall M. Data Mining: Practical Machine Learning Tools and Techniques. Burlington: Morgan Kaufmann; 2011.

12- Guh RS, Wu TCJ, Weng SP. Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes. Expert Syst Appl. 2011;38(4):4437-49.

13- Jurisica I, Mylopoulos J, Glasgow J, Shapiro H, Casper RF. Case-based reasoning in IVF: Prediction and knowledge mining. Artif Intell Med. 1998;12(1):1-24.

14- Cai Qf, Wan F, Huang R, Zhang HW. Factors predicting the cumulative outcome of IVF/ICSI treatment: A multivariable analysis of 2450 patients. Hum Reprod. 2011;26(9):2532-40.