

## به کارگیری خوشبندی فازی در ریزآرایه DNA

مسنونامدی<sup>\*</sup>، دکتر محمد علوی‌مجد<sup>۱</sup>، دکتر یداد... محمدابی<sup>۲</sup>، بهار نقوی<sup>۳</sup>

۱. کارشناس ارشد آمار زیستی، مرکز تحقیقات بیماریهای گوارش و کبد، دانشگاه علوم پزشکی شهید بهشتی
۲. دانشیار، گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی
۳. استاد، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی
۴. مری، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی شهید بهشتی

### چکیده

سابقه و هدف: فناوری ریزآرایه برای بررسی همزمان بیان هزاران ژن در بازه وسیعی از ژنومیک، نظریه شناسایی ژنهای، اکتشاف داروها و تشخیص‌های کلینیکی مورد استفاده قرار گرفته است. آزمایشها ای که بر اساس فناوری ریزآرایه انجام می‌شوند حجم بسیار زیادی از داده‌ها را فراهم می‌کنند که در مطالعات بیولوژیک بی‌نظیر بوده است. نرمال سازی، خوشبندی، طبقه‌بندی و ... از جمله روش‌های مورد استفاده در تحلیل آماری این نوع داده‌هاست. هدف این مقاله بررسی نحوه به کارگیری روش خوشبندی فازی در داده‌های ریزآرایه DNA است.

مواد و روش‌ها: تحقیق به روش توصیفی انجام شده و داده‌های بیان ژنی سرطان خون گلوب و همکاران (۱۹۹۹) که بر اساس روش آرایه الیگونوکلئوتید تولید شده و از طریق اینترنت در اختیار عموم قرار دارد، با استفاده از روش آماری خوشبندی فازی، مورد تجزیه و تحلیل قرار گرفته است. خوشبندی داده‌های ریزآرایه به صورت خیلی محدود و گرایش بیشتر به سمت خوشبندی کلاسیک در ایران صورت پذیرفته است و این مقاله شروعی در خوشبندی فازی داده‌های ریزآرایه به شمار می‌آید. مجموعه داده‌ها شامل ۲۰ بیمار مبتلا به سرطان خون لنفوئیدی حاد (ALL) و ۱۴ بیمار مبتلا به سرطان خون میلوئیدی حاد (AML) است. کارایی روش خوشبندی فازی با توجه به گروه‌بندی واقعی نمونه‌ها (ALL و AML) مورد ارزیابی قرار گرفت. نرمافزار R برای تحلیل داده‌ها استفاده شد.

یافته‌ها: ویژگی روش خوشبندی فازی در تشخیص افراد AML، %۹۰ و حساسیت آن %۹۳ و صحت آن %۹۱ به دست آمد که نشان‌دهنده عملکرد خوب این روش است. نمونه سی و یکم که بر اساس یافته‌های بالینی در گروه ALL قرار دارد در گروه AML قرار گرفت، همچنانی نمونه‌های دوم و هفدهم که بر اساس یافته‌های بالینی در گروه ALL قرار دارد طبق نتایج در گروه AML قرار گرفتند که از نظر بالینی می‌توانند قابل توجه باشند.

نتیجه‌گیری: خوشبندی فازی اطلاعات نسبتاً قابل قبولی درباره ساختار داده‌ها فراهم می‌کند که با توجه به انطباق نتایج این روش با گروه‌بندی واقعی داده‌ها، از این روش آماری می‌توان در مواردی که اطلاع دقیقی درباره گروه‌بندی واقعی داده‌ها در دست نیست، استفاده کرد. به علاوه با بررسی نتایج خوشبندی ممکن است زیرگروه‌هایی از نمونه‌ها را به نحوی تمایز کرد که برای انطباق آن با یافته‌های بالینی، پژوهش‌های آزمایشگاهی یا بالینی جدیدی لازم باشد.

**واژگان کلیدی:** ریزآرایه DNA، بیان ژن، خوشبندی کلاسیک، خوشبندی فازی، سرطان خون

سلول در شرایط فیزیولوژیک و پاتولوژیک را تغییر داده‌اند.  
فناوری ریزآرایه برای بررسی همزمان بیان هزاران ژن در بازه وسیعی از ژنومیک، نظریه شناسایی ژنهای، اکتشاف داروها و تشخیص‌های کلینیکی به صورت موفقیت‌آمیزی مورد استفاده قرار گرفته است (۱).

داده‌های سطوح بیان ژنهای اطلاعات ارزشمندی در مورد شبکه‌های بیولوژیک، حالات سلولی و فهمیدن کارکرد ژنهای در

### مقدمه

در سالهای اخیر فناوری ریزآرایه، امکان کنترل بیان هزاران ژن را به صورت همزمان فراهم کرده و ژنومیکس و پروتئومیکس اساس شیوه‌های علمی مطالعه پایه مولکولی رفتارهای بافت و

\*نویسنده مسئول مکاتبات: محسن واحدی، تهران، اوین، خیابان تابناک، پیمارستان آیت الله طالقانی، مرکز تحقیقات بیماریهای گوارش و کبد؛ پست الکترونیک: mohsenvahedi540@gmail.com

(Fuzzy C-Means) نمی‌دهند، روش خوشبندی فازی FCM می‌تواند گره‌گشا باشد (۷). همچنین با مقایسه اعتمادپذیری تحلیل‌های ریزآرایه با توجه به دو روش FCM و مدل‌های آمیخته نرمال (Normal Mixture Modeling) در حالت مدل‌های آمیخته، با توجه به سرعت بیشتر FCM در چنین تحلیل‌هایی این روش برتر از NMM می‌باشد (۸).

خوشبندی فازی را می‌توان به عنوان حالت تعمیم یافته افزار قطعی (Hard Partitioning) در نظر گرفت. در یک افزار قطعی (خوشبندی سلسله مراتبی متعلق به این ردۀ می‌باشد) هر عنصر فقط به یک خوشبندی تعلق دارد.

در روش خوشبندی فازی برای هر عنصر میزان قرار گرفتن در خوشبندی‌های مختلف بوسیله ضرایب عضویت (Membership Coefficients) که اعدادی در بازه ۰ تا ۱ است سنجیده می‌شود. ضریب عضویت عنصر در هر خوشبندی که بیشتر بود عنصر متعلق به آن خوشبندی است.

اصلًا پایه ریزی الگوریتم خوشبندی فازی برای فائق آمدن بر وضعیت بروز مشکل همپوشی خوشبندی‌ها که از طریق وجود عنصرهایی که در تعلق آنها به خوشبندی‌های مختلف ابهام وجود دارد ابداع شد. اساس آن مبتنی بر تقسیم  $n$  آزمودنی به  $c$  خوشبندی از طریق محاسبه و تعیین ضریب عضویت هر آزمودنی به هر خوشبندی می‌باشد.

امتیاز اصلی خوشبندی فازی نسبت به خوشبندی قطعی ارایه اطلاعات دقیق‌تری درباره ساختار داده‌ها می‌باشد. از طرف دیگر این ممکن است به عنوان یک اشکال در نظر گرفته شود، زیرا مقدار خروجی با افزایش تعداد عناصر و تعداد خوشبندی‌ها به سرعت زیاد می‌شود که در نتیجه زمان زیادی برای تحلیل صرف می‌شود. برای بررسی داده‌های ریزآرایه DNA می‌توان از نرم‌افزارهای آماری نظیر SAS، S-plus، STATA و R استفاده کرد. R به علت توانایی بالا در کار کردن با داده‌های حجمی رواج بیشتری دارد (۹).

هدف این مقاله بررسی نحوه به کارگیری روش خوشبندی فازی در داده‌های ریزآرایه DNA است.

## مواد و روش‌ها

این مطالعه از نوع مشاهده‌ای- مقطعي می‌باشد. فناوری ریزآرایه DNA و روش‌های تحلیل داده‌های آن جدید می‌باشند، بنابراین در ایران و خیلی از کشورهای دیگر تحقیقاتی در زمینه تولید این گونه داده‌ها صورت نگرفته است، لذا اغلب محققان ناچارند که از داده‌های بانک‌های اطلاعاتی اینترنتی نظیر GenBank استفاده کنند (۱۰).

بر دارد. یک هدف از تحلیل داده‌های بیان ژن، تعیین چگونگی تأثیر بیان ہر ژن منفرد روی بیان ژنهای دیگر در همان شبکه ژنتیکی است. هدف دیگر، مشخص کردن این نکته است که چگونه ژنهای در سلولهای سالم و بیمار بیان می‌شوند. کاربرد علمی بررسی بیان ژن ریزآرایه مدیریت و کنترل سرطان و بیماریهای عفونی است. هدف اصلی این مطالعات، تعیین و شناسایی فرایند پاتولوژیک مرتبط با نوع بیماری و مرحله آن و نیز پیش‌بینی پاسخ به درمان خاصی است. همچنین برخی از مسایل در زمینه تشخیص، با استفاده از تحلیل داده‌های بیان ژن قابل حل شده‌اند (۲).

ریزآرایه ابزاری برای اندازه‌گیری و کسب اطلاعات از بیان ژنهاست. هر توالی ژنی شناخته شده مورد نظر به عنوان یک پربو (Prob) روی یک آرایه (Array) شیشه‌ای یا نایلوونی چاپ می‌شود. mRNA از بافت یا نمونه خون با رنگهای فلورسنت علامت‌گذاری می‌شود و پروبها بر روی یک آرایه هیبرید می‌شوند. دو نوع آرایه بیشترین کاربرد را دارند: ۱- آرایه‌های بر پایه DNA مکمل (DNA Complementary Spotted) ۲- آرایه‌های الیگونوکلئوتید (Oligonucleotide array) که به اختصار الیگو گفته می‌شود (۳).

اغلب داده‌های حاصل از این دو روش در ماتریس بیان ژنی (Gene Expression) ذخیره می‌شوند که سطوحی آن ژنهای و ستونهای آن افراد نمونه می‌باشند.

آنالیز پایه داده‌های بیان ژن شامل یک مرحله پیش پردازش و آماده‌سازی مجموعه داده برای مراحل آنالیز سطوح بالاتر است. پیش پردازش داده‌های خام می‌تواند اثرات عمیقی روی مراحل بعدی آنالیز داشته باشد. استفاده از روش‌های آماری برای تحلیل داده‌های حاصل از این فناوری می‌تواند گام مؤثری در جهت تشخیص و درمان بیماریها داشته باشد. یکی از روش‌های آماری که در تحلیل این داده‌ها به صورت فرایندی مورد استفاده قرار می‌گیرد خوشبندی است. از سال ۱۹۹۸ خوشبندی داده‌های بیان ژنی شروع گردیده است (۴). در ابتدا فقط روش‌های خوشبندی کلاسیک نظیر خوشبندی سلسله مراتبی مورد استفاده قرار می‌گرفت که در بسیاری از موارد روش‌های کارا و مفید بودند ولی در بعضی از موقع کاستی‌هایی نیز داشتند که باعث شد تا روش‌های خوشبندی غیرکلاسیک نظیر خوشبندی فازی در نظر گرفته شود؛ مثلاً به منظور از بین بردن مشکل خوشبندی اشتباه داده‌های ریزآرایه از روش خوشبندی فازی استفاده شد (۵). متداول‌تر استفاده از این روش نیز ارایه گردید (۶). در مواردی که ساختار داده‌ها پیچیده است و خوشبندی‌های رایج جواب

که نشان‌دهنده این نکته است که عضویت‌ها نمی‌تواند منفی باشد و مجموع ضرایب عضویت یک عنصر روی خوشبندی برابر با یک است. از تابع لاگرانژ که برای بهینه کردن (#) استفاده کردیم.

به خاطر محدودیت‌هایی که برای مورد‌های فازی است راه حل‌های خوشبندی قطعی مورد نظر قرار می‌گیرد. میزان تفاوت یک راه حل فازی از خوشبندی قطعی را می‌توان از ضریب افزایش دان (Dunn's Partition Coefficient) ارزیابی کرد که به صورت مجموع مربعات همه ضرایب عضویت تقسیم بر تعداد عناصر به دست می‌آید.

$$F_k(U) = \sum_{i=1}^n \sum_{v=1}^k u_{iv}^2 / n$$

که  $U$  ماتریس ضرایب عضویت است.

$$U = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1k} \\ u_{21} & u_{22} & \dots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \dots & u_{nk} \end{bmatrix}$$

برای یک افزایش  $u_{iv}$  محدود شده به صفر (یک) ( $F_k(U)$ ) ماکزیمم مقدار یک را می‌گیرد در حالی که مینیمم مقدار  $1/k$  را وقتی همه  $u_{iv} = 1/k$  است به خود اختصاص می‌دهد. این ضریب می‌تواند به صورتی که از ۱ (خوشبندی قطعی) تا ۰ (تماماً فازی) تغییر کند، نرمال شود. این کار مستقل از تعداد خوشبندی‌ها به وسیله تبدیل انجام شود:

$$F'_k(U) = \frac{F_k(U) - (1/k)}{1 - (1/k)} = \frac{kF_k(U) - 1}{k - 1}$$

این ضریب نرمال شده اغلب به عنوان شاخص غیرفازی (Nonfuzziness Index) نامیده می‌شود (۱۳).

اگر بخواهیم یک خوشبندی فازی را به خوشبندی قطعی تبدیل کنیم باید  $w_{iq} = 1$  برای خوشبندی  $q$  که بزرگترین مقدارهای  $u_{iv}$  را دارد تعریف کنیم. در مواردی که گره وجود دارد به صورت انتخابی عمل می‌کنیم. این تبدیل برای مقایسه فازی با یک راه حل قطعی مورد استفاده قرار می‌گیرد.

برای مقایسه نتایج دو روش فازی و غیرفازی و بررسی این موضوع که نتایج آنها تا چه حد با هم تطابق دارند، لازم است نتایج حاصل از روش غیرفازی با مقادیر و ضرایب عضویت عناصر به خوشبندی متناظر حاصل از روش فازی مقایسه شود که بالا بودن مقدار ضریب عضویت برای یک عنصر مثل  $i$  به خوشبندی که مثل  $v$  است همین عنصر بر طبق روش غیرفازی در آن قرار گرفته است بیانگر مطابقت نتایج دو روش است.

در این تحقیق از داده‌های سرطان خون که توسط گلوب و همکاران (Golub) در سال ۱۹۹۹ منتشر یافته استفاده شد (۱۱). نمونه‌های سرطان خون شامل ۲۴ نمونه مغز استخوان و ۱۰ نمونه خون می‌باشد که همگی در زمان تشخیص سرطان خون گرفته شده‌اند. ۲۰ نمونه از بیماران با سرطان خون حاد لنفوییدی (ALL) و ۱۴ نمونه از بیماران با سرطان خون حاد میلوییدی (AML) می‌باشند، که به صورت نمونه‌گیری مبتنی بر هدف و غیرتصادفی انتخاب شده‌اند و بر اساس روش آرایه‌ای لیگنوکلئوتید بیان ژنها به دست آمده است. این داده‌ها از طریق اینترنت در اختیار عموم قرار دارد (۱۲).

در این داده‌ها هم نظری پیشتر داده‌های مطالعات بیان ژنی، داده‌ها چوله و دارای نقاط پرت بودند، لذا لازم بود ابتدا یک پیش‌پردازش بر روی داده‌ها صورت بگیرد.

با توجه به توصیه‌های گلوب و همکاران (۱۹۹۹) موارد ذیل برای پیش‌پردازش داده‌ها در نظر گرفته شد:

(۱) انتخاب حد آستانه برای داده‌ها: حداقل مقدار هر داده و حداقل مقدار ۱۶۰۰۰ باشد، یعنی داده‌هایی که مقدار بیان ژنی آنها کمتر از ۱۰۰ بود را ۱۰۰ منظور گردید و داده‌هایی که بیشتر از ۱۶۰۰۰ بودند را ۱۶۰۰۰ گرفته شد.

(۲) فیلتر کردن: خارج کردن داده‌هایی که  $5 \leq \text{max}/\text{min} \leq 500$  و  $\text{min} \leq \text{max} - \text{min} \leq 500$  بودند. منظور از  $\text{max}$  و  $\text{min}$  به ترتیب حداقل و حداقل سطوح بیان یک ژن خاص در طول ۳۴ نمونه می‌باشد.

(۳) انجام تبدیل لگاریتمی بر روی داده‌ها

(۴) استاندارد کردن: تبدیل نرمال استاندارد بر روی داده‌ها زده شد در نتیجه برای هر نمونه سطوح بیان ژنی دارای میانگین صفر و واریانس یک شد.

با انجام این پیش‌پردازش تنها ۲۹۱۷ ژن در ماتریس داده‌ها باقی ماندند و ماتریس بیان ژنی  $34 \times 2917$  برای تحلیل آماری مورد استفاده قرار گرفت.

تکنیک خوشبندی فازی را با مینیمم کردن تابع هدف زیر به دست آورده‌ایم:

$$C = \sum_{v=1}^k \frac{\sum_{i,j=1}^n u_{iv}^2 u_{jv}^2 d(i, j)}{2 \sum_{j=1}^n u_{jv}^2} \#$$

$d(i, j)$  نشان‌دهنده تفاوت یا فاصله بین دو عنصر  $i$  و  $j$  است،  $u_{iv}$  ضریب عضویت نامعلوم عنصر  $i$  به خوشبندی  $v$  است. توابع عضویت به صورت زیر مقید شده‌اند:

$$u_{iv}^2 \geq 0 \quad \text{for } i = 1, \dots, n ; v = 1, \dots, k$$

$$\sum_v u_{iv} = 1 \quad \text{for } i = 1, \dots, n$$

**جدول ۱- انجام خوشبندی فازی به دو خوشه و مقادیر سایه نما**

نحوه نه	نوع لوسمی	ضرایب عضویت		نردهای خوشه	مقادیر سایه نما
		خوشه ۱	خوشه ۲		
۱	ALL B-cell	۰/۷۲۷	۰/۲۷۳	۱	۰/۴۶۰
۲	ALL B-cell	۰/۴۵۱	۰/۵۴۹	۲	۰/۰۴۱
۳	ALL B-cell	۰/۷۸۸	۰/۲۱۲	۱	۰/۴۵۸
۴	ALL B-cell	۰/۷۸۲	۰/۲۱۸	۱	۰/۴۳۲
۵	ALL B-cell	۰/۹۰۱	۰/۰۹۹	۱	۰/۴۱۴
۶	ALL B-cell	۰/۵۰۹	۰/۴۹۱	۱	۰/۳۹۵
۷	ALL B-cell	۰/۸۸۲	۰/۱۱۸	۱	۰/۳۸۴
۸	ALL B-cell	۰/۷۸۵	۰/۲۱۵	۱	۰/۳۴۵
۹	ALL B-cell	۰/۸۹۹	۰/۱۰۱	۱	۰/۳۳۹
۱۰	ALL B-cell	۰/۸۷۵	۰/۱۲۵	۱	۰/۳۲۴
۱۱	ALL B-cell	۰/۹۱۹	۰/۰۸۱	۱	۰/۳۱۰
۱۲	ALL B-cell	۰/۸۱۶	۰/۱۸۴	۱	۰/۳۰۹
۱۳	ALL B-cell	۰/۷۳۸	۰/۲۶۲	۱	۰/۲۹۱
۱۴	ALL B-cell	۰/۸۰۸	۰/۱۹۲	۱	۰/۲۷۱
۱۵	ALL B-cell	۰/۸۲۵	۰/۱۷۵	۱	۰/۲۶۷
۱۶	ALL B-cell	۰/۸۹۷	۰/۱۰۳	۱	۰/۲۵۳
۱۷	ALL T-cell	۰/۴۳۵	۰/۵۶۵	۲	۰/۰۲۱
۱۸	ALL B-cell	۰/۵۲۸	۰/۴۷۲	۱	۰/۲۲۰
۱۹	ALL B-cell	۰/۸۱۸	۰/۳۸۲	۱	۰/۱۲۸
۲۰	ALL B-cell	۰/۷۳۱	۰/۲۶۹	۱	۰/۱۰۸
۲۱	AML	۰/۱۷۸	۰/۸۲۲	۲	۰/۳۵۸
۲۲	AML	۰/۱۶۵	۰/۸۳۵	۲	۰/۳۳۱
۲۳	AML	۰/۱۳۴	۰/۸۶۴	۲	۰/۳۲۹
۲۴	AML	۰/۱۲۱	۰/۸۷۹	۲	۰/۳۱۴
۲۵	AML	۰/۲۵۸	۰/۷۴۲	۲	۰/۲۸۶
۲۶	AML	۰/۱۳۸	۰/۸۶۲	۲	۰/۲۸۶
۲۷	AML	۰/۱۳۶	۰/۸۶۴	۲	۰/۲۶۵
۲۸	AML	۰/۲۶۶	۰/۷۳۴	۲	۰/۲۶۳
۲۹	AML	۰/۳۴۸	۰/۷۵۲	۲	۰/۲۵۶
۳۰	AML	۰/۱۴۵	۰/۸۵۵	۲	۰/۲۴۵
۳۱	AML	۰/۶۸۵	۰/۳۱۵	۱	۰/۲۰۲
۳۲	AML	۰/۱۱۷	۰/۸۸۳	۲	۰/۱۲۶
۳۳	AML	۰/۲۰۶	۰/۷۹۴	۲	۰/۱۱۳
۳۴	AML	۰/۱۴۹	۰/۸۵۱	۲	۰/۱۰۸

نمودار ۱ نمونه‌ها را در دو خوشه با روش خوشبندی فازی نشان می‌دهد. با توجه به نمودار ۱ نتایج انتساب نمونه‌ها به دو خوشه در جدول ۲ آمده است. با توجه به این جدول ویژگی روش خوشبندی فازی در تشخیص افراد AML،  $90\%$  حساسیت  $93\%$  و صحت  $91\%$  به دست آمد.

نمودار ۱ بیانگر عدم فشردگی خوشه‌ها می‌باشد، برای ارزیابی کارایی روش خوشبندی فازی از نمودار سایه‌نما استفاده کردیم. برای رسم این نمودار ابتدا مقدارهای سایه‌نما را به دست آوردیم، سپس با توجه به مقادیر سایه‌نما، نمودار سایه‌نما را که در نمودار ۲ به دست آوردهیم که ضریب سایه‌نما  $0/۳۱$  به دست آمد، لذا ساختار خوشه‌ای ضعیف است و توصیه می‌شود از الگوریتم دیگری استفاده شود. برای بررسی بیشتر خوشبندی فازی در سه خوشه نیز انجام گردید که نتیجه آن در نمودار ۳ آمده است که با توجه به آن خوشه اول و دوم

اجرای روش خوشبندی فازی با استفاده از نرم افزار R دارای عبارات و مفاهیمی است. یکی از این مفاهیم نمودار سایه‌نما (Silhouette) است. نمودار سایه‌نما اولین بار توسط رسیو معرفی گردید (۱۴). در این نمودار هر خوشه به صورت یک سایه‌نما مطرح می‌شود که در آن وضعیت هر آزمودنی که در آن خوشه قرار گرفته است، به لحاظ شدت تعلق به آن خوشه مشخص می‌باشد. کل خوشبندی با قرار گرفتن چند سایه‌نما (به تعداد خوشه‌های تعیین شده) در کنار هم معین می‌گردد و به کاربر این اجازه را می‌دهد که به مقایسه کیفیت خوشه‌های تشکیل شده بپردازد. چنین نمودار سایه‌نمای در هنگامی که مقیاس داده‌ها نسبی است، بسیار مفید است، با استفاده از آن به راحتی فشرده بودن داخلی هر خوشه و تفکیک خوشه‌های مختلف می‌تواند مورد بررسی قرار گیرد.

برای هر خوشه میانگین مقادیر سایه‌نما را به دست می‌آوریم. بزرگترین مقدار میانگین در بین خوشه‌های مختلف به عنوان ضریب سایه‌نما نامیده می‌شود که به صورت جدول زیر تفسیر می‌شود (۱۳).

ضریب سایه‌نما	تفسیر مربوطه	مقدار عددی
۰/۷۱-۱	ساختار خوشه‌ای بسیار مناسبی تشکیل شده است	۰/۷۱
۰/۵۱-۰/۷۰	ساختار خوشه‌ای معقولی تشکیل شده است.	۰/۵۱
۰/۲۶-۰/۵۰	ساختار خوشه‌ای ضعیف است و توصیه می‌شود از الگوریتم دیگری استفاده شود.	۰/۲۶
۰-۰/۲۵	اصولاً یافتن ساختار مناسب برای این داده‌ها مشکل می‌باشد.	۰-۰/۲۵

در این تحقیق داده‌های بیان ژنی پس از پیش‌پردازش اولیه با روش خوشبندی فازی، خوشبندی گردید و برای ارزیابی خوشبندی از مقادیر و نمودار سایه‌نما استفاده شد.

**یافته‌ها**

روش خوشبندی فازی در مورد وضعیت‌های مبهمی که در داده‌ها رخ می‌دهد یعنی حالتی که خوشه‌ها از طریق وجود عنصرهایی که از لحاظ تعلق به خوشه‌های مختلف در یک وضعیت بینایی‌نی قرار می‌گیرند و حالت همپوشی در آنها ایجاد می‌شود، کاربرد دارد. خوشبندی فازی را ابتدا با ۲ خوشه شروع کردیم، سپس ضریب عضویت هر یک از نمونه‌ها به هر خوشه محاسبه گردید. بر اساس این که ضرایب عضویت در کدام خوشه بیشتر باشد می‌توان نمونه را به آن خوشه اختصاص داد، نتایج این اختصاص در جدول ۱ آمده است.

برای بیان میزان تفاوت یک راه حل فازی از خوشبندی قطعی از ضریب افزایش دان استفاده کردیم که مقدار آن  $0/۸۷۴$  و شاخص غیرفازی  $0/۷۴۸$  به دست آمد.

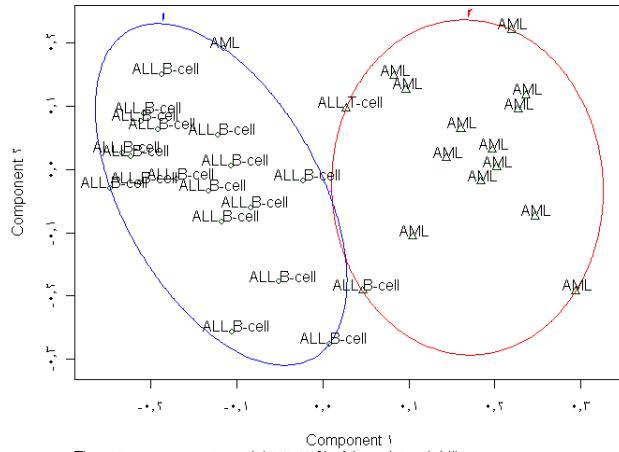
## به کارگیری خوشبندی فازی در ریزآرایه

**نمودار ۳- نمایش نمونه‌ها در سه خوشیده در روش خوشبندی فازی**

هر چقدر که تعداد خوشه‌ها را زیادتر کنیم، امکان یافتن زیرگروه‌های جدید افزایش می‌یابد، مثلاً در تعداد سه خوشه که در نمودار ۳ نشان داده شده است اعضای خوشه اول همگی ALL هستند. شاخص غیرفازی ۰/۷۴۸ به دست آمد که برتری روشهای افزار قطعی را نسبت به روش فازی در این داده‌ها نشان می‌دهد. ضریب سایه‌نما ۰/۳۱ این نظر را تأیید می‌کند، که در این داده‌های عملکرد روشهای دیگر خوشه‌بندی از فازی بهتر است، شاید دلیل آن عدم وجود موارد بینایی در داده‌ها است. در مواردی که در داده‌ها موارد بینایی وجود دارد روش فازی مناسب‌تر است چون امکان انتخاب را به پژوهشگر می‌دهد تا در مورد قرار گرفتن نمونه‌ها در خوشه‌ها تصمیم‌گیری کند. بر طبق نتایج خوشه‌بندی نمونه‌ها سی و یکم که بر اساس یافته‌های بالینی در گروه AML قرار دارد، در گروه ALL قرار گرفت، همچنین نمونه‌های دوم و هفدهم که بر اساس یافته‌های بالینی در گروه ALL قرار دارد طبق نتایج در گروه AML قرار گرفتند که از نظر بالینی می‌توانند قابل توجه باشند. نتایج روش خوشه‌بندی فازی با روشهای دیگر خوشه‌بندی که تاکنون بر روی این داده‌ها انجام شده است، همخوانی دارد (۱۵) ولی خوشه‌بندی فازی این برتری را دارد که محقق با بررسی ضرایب عضویت نمونه‌ها خصوصاً نمونه‌هایی که با یافته‌هایی بالینی مغایرت دارد تصمیم بهتری را بگیرد (۱۶ و ۱۷). همچنین با بررسی تعدادی از مطالعات بیان ژنی که از روش خوشه‌بندی فازی FCM استفاده کرده‌اند، می‌بینیم که از این روش می‌توان به منظور از بین بردن مشکل خوشه‌بندی اشتباه داده‌های ریزآرایه و مواردی که ساختار داده‌ها پیچیده است و خوشه‌بندی‌های رایج جواب نمی‌دهند، استفاده کرد (۱۷ و ۱۸). همچنین روش FCM د، تحلیا های، بآ، ابه د، حالت مدل‌های، آمخته

مربوط به ALL و خوش سوم AML است. ضریب افزای دان و شاخص غیرفازی  $0/674$  به دست آمد.

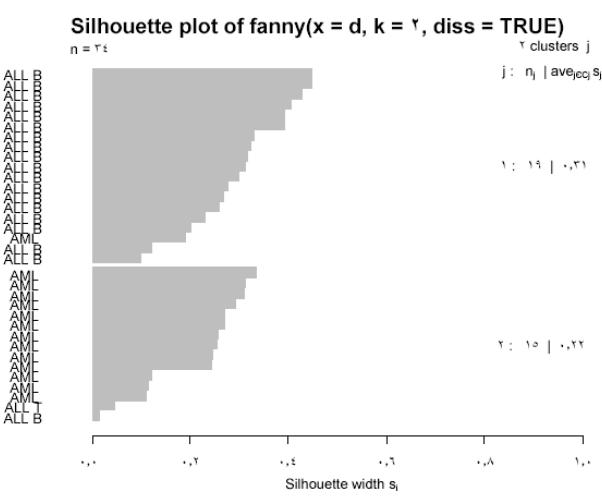
```
clusplot(fanny(x = d, k = 4, diss = TRUE, memb.exp = 1.2, maxit = 100))
```



These two components explain 78.41% of the point variability.

## جدول ۲- انتساب بیماران با روش خوشبندی فازی به دو خوش

لوسومی	خوش بندی	خوشه ۱	خوشه ۲	جمع
AML	ALL	۱۸	۲	۲۰
حساسیت	ویرگی	۱	۱۳	۱۴
%۹۳	%۹۰	٪۹۱	٪۷۶	٪۷۵



نمودا، ۲- نمودا، سایه‌نما د، دو خمشه به، وش، خمشه‌بندی، فاز،

بحث

با توجه به نتایج روش خوشبندی فازی در انتساب نمونه‌ها به دو خوشه توانایی کامل دارد زیرا از ۲۰ نمونه ALL ۱۸ نمونه و از ۱۴ نمونه AML ۱۳ نمونه را به درستی اختصاص داده است.

شده‌اند به ارزیابی این تجربیات قبل از به کار بردن آنها در بیماران پرداخت. از طرفی روش‌های پیش‌بینی طبقات را می‌توان جهت تشخیص نتایج بالینی در آینده، از قبیل واکنش به یک دارو یا میزان بقای افراد بیمار را نیز به کار برد (۱۹).

فناوری تولید داده‌های ریزآرایه در ایران و بسیاری از کشورها هنوز مورد استفاده قرار نگرفته است و این نوع داده‌ها تنها در کشورهای معینی تولید می‌شود و پژوهشگرانی نظری نویسندگان مقاله ناگزیر به استفاده از داده‌هایی هستند که توسط دیگران جمع‌آوری شده است، بنابراین کاستیهای احتمالی در صحت داده‌ها از محدودیت این تحقیق به شمار می‌آید.

### نتیجه‌گیری

خوشبندی فازی اطلاعات نسبتاً قابل قبولی درباره ساختار داده‌ها فراهم می‌کند که با توجه به انطباق نتایج این روش با گروه‌بندی واقعی داده‌ها، از این روش آماری می‌توان در مواردی که اطلاع دقیقی درباره گروه‌بندی واقعی داده‌ها در دست نیست، استفاده کرد. به علاوه با برسی نتایج خوشبندی ممکن است زیرگروه‌هایی از نمونه‌ها را به نحوی متمایز کرد که برای انطباق آن با یافته‌های بالینی، پژوهش‌های آزمایشگاهی یا بالینی جدیدی لازم باشد.

### تشکر و قدردانی

این مقاله حاصل طرح تحقیقاتی است که اعتبار آن توسط دانشکده پیراپزشکی دانشگاه علوم پزشکی شهید بهشتی تأمین شده است که در اینجا از معاونت محترم آموزشی پژوهشی دانشکده سپاسگزاری می‌شود.

اعتمادپذیری و سهولت انجام بیشتری نسبت به مدل‌های آمیخته دارد (۸). با توجه به این موارد استفاده از روش خوشبندی فازی روز به روز افزونتر می‌گردد. نتایج به دست آمده از ریزآرایه منجر به ایجاد دیدگاه‌های تازه‌ای در مورد نحوه شکل‌گیری، پیشرفت و پاسخ به درمان بیماران سلطانی گردیده است. با توجه به اینکه توالی کامل ژنوم انسانی در دسترس می‌باشد بررسی کامل نسخه‌برداری در سلولهای نرمال و سلطانی امکان‌پذیر گردیده است و همراه با تکامل همزمان ابزارهای ضروری انفورماتیک و آنالیز داده‌ها جهت تبدیل و تفسیر آنها، نحوه نگرش به سلطان دچار تحول شگرفی گردیده است. ترکیبی از روش‌های ژنومیک و پروتئومیکس احتمالاً سبب پیشرفت‌های عمیقتری در این زمینه خواهد شد (۱۸).

فرضیات ویژه و نهایی در مطالعات مختلف می‌توانند بسیار متنوع باشند به همین علت در انتخاب روش‌های آماری مناسب جهت مطالعه افراد باید دقت زیادی را صرف نمود. در واقع روش‌های خوشبندی برای مشخص کردن زیرگروه‌های احتمالی در مورد هر نوع سرطان دیگر را نیز به کار برد. روش‌های کشف طبقات را می‌توان همچنین جهت تحقیق پیرامون مکانیزم‌های اساسی که باعث تشخیص انواع سرطانها می‌شود به کار بست. به عنوان مثال، جهت مطالعه روند و نحوه بیان ژنها می‌توان سرطانهای متفاوتی را با هم در یک مجموعه واحد ترکیب نمود و بعد از حذف ژنهایی که با نوع بافت‌های مورد نظر همبستگی بالایی دارند، نمونه‌ها را بر اساس ژنهای باقیمانده خوشبندی کرد.

این روشها را می‌توان در کنار روش‌های بالینی به کار برد تا به اطمینان بیشتری به تشخیص بیماری دست یافت. در حقیقت این فناوری فرصتی را جهت افزایش دقت تجربیات بالینی فراهم کرده که می‌توان در سرطانهایی که به خوبی مطالعه

### REFERENCES

- Beltrame F, Papadimitropoulos A, Porro I, Scaglione S, Schenone A, Torterolo L, et al. GEMMA-A Grid environment for microarray management and analysis in bone marrow stem cells experiments. Future Generation Computer Systems 2007;23(3):382-90.
- Ho SY, Hsieh CH, Chen HM, Huang HL. Interpretable gene expression classifier with an accurate and compact fuzzy rule base for microarray data analysis. BioSystems 2006;85(3):165-76.
- Satagopan JM, Panageas KS. A statistical perspective on gene expression data analysis. Stat Med 2003;22(3):481-99.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95(25):14863-8.
- Woolf PJ, Wang Y. A fuzzy logic approach to analyzing gene expression data. Physiol Genomics 2000;3(1):9-15.
- Inza I, Sierra B, Blanco R, Larrañaga P. Gene selection by sequential search wrapper approaches in microarray cancer class prediction. Journal of Intelligent and Fuzzy Systems 2002;12(1):25-34.
- Dembele D, Kastner P. Fuzzy C-means method for clustering microarray data. Bioinformatics 2003;19(8):973-80.

8. Asyali MH, Alci M. Reliability analysis of microarray data using fuzzy c-means and normal mixture modeling based classification methods. *Bioinformatics* 2005;21(5):644–49.
9. The R Project for Statistical Computing. Available from: URL:<http://www.r-project.org>.
10. National Center for Biotechnology Information. Available from: URL: <http://www.ncbi.nlm.nih.gov>.
11. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531–7.
12. Cancer Program Data Sets. Available from: URL: <http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi>.
13. Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to cluster Analysis*. Wiley: New York, 1990.
14. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational Application Mathematics* 1987;20:53-65.
15. Vahedi M, Alavi Majd H, Mehrabi Y, Naghavi B. Gene expression data clustering and it's application in differential analysis of leukemia. *Journal of Semnan University of Medical Sciences* 2008;9(2):163-9. (Full Text in Persian)
16. Alavimajd H, Vahedi M, Mehrabi Y, Naghavi B. Clustering approach in DNA microarray analysis. *Pejouhesh dar pezeshki* 2007;31(1):19-26. (Full Text in Persian)
17. Vahedi M. Application of clustering methods in gene expression data [Dissertation]. Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2006.
18. Korenberg MJ. *Microarray Data Analysis: Methods and Applications (Methods in Molecular Biology)*. New Jersey: Humana Press; 2007.
19. Knudsen S. *Cancer Diagnostics with DNA Microarrays*. New Jersey: Wiley; 2006.