

تعیین روش درمانی بیماران مبتلا به سرطان پستان به کمک روش‌های هوش مصنوعی

ستایش صادقی^۱، عباس رضایی^۲، امین گلاب پور^۳

^۱ دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد کرمان، کرمان، ایران.

^۲ مربی دانشگاه آزاد اسلامی واحد کرمان، کرمان، ایران.

^۳ استادیار دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

نام نویسنده مسئول:

امین گلاب پور

چکیده

امروزه در دنیای پزشکی شاهد جمع آوری داده‌های فراوان در مورد بیماری‌های مختلف هستیم. تحقیق بر روی این داده‌ها و بدست آوردن نتایج و الگوهای مفید در رابطه با بیماری‌ها یکی از اهداف استفاده از این داده‌هاست. حجم زیاد داده‌ها و سردرگمی حاصل از آن مشکلی است که مانع رسیدن به نتایج قابل توجه می‌شود. در اینجاست که نیاز به یک سیستم مکانیزه برای کمک به کشف الگوهای موجود و همچنین پیش‌بینی و ارائه روش درمانی بیماری‌ها کاملاً احساس می‌شود. دانش داده‌کاوی به عنوان ارائه کننده این سیستم مکانیزه کمک شایانی در پیشرفت‌های پزشکی، به ویژه در زمینه تشخیص و الگوهای درمانی بیماری‌های گوناگون کرده است. از جمله این بیماری‌های می‌توان به سرطان پستان اشاره کرد که شیوه فراوانی داشته و در صورت درمان نامناسب به عوارض جدی و خطرناک و حتی مرگ شود. لذا انتخاب راه حل درمانی درست ضروری است. در این مقاله پس معرفی سرطان پستان، مروری بر انواع روش‌های درمانی تغذیه در شیمی درمانی خواهیم داشت. سپس به کمک درخت تصمیم برای انتخاب بهترین روش تغذیه در شیمی درمانی ارائه می‌دهیم.

واژگان کلیدی: سرطان پستان - هوش مصنوعی - درخت تصمیم.

مقدمه

سرطان پستان شایع‌ترین سرطان زنان در امریکا و دومین علت مرگ در اثر سرطان در این کشور است (۱،۲). بر اساس آمار سازمان بهداشت جهانی در سال ۲۰۱۱ سرطان پستان پنجمین علت مرگ ناشی از سرطان در جهان با حدود ۵۰۸،۰۰۰ مورد مرگ از بین حدود ۵۴ میلیون مرگ در جهان می‌باشد. این سرطان در بین زنان، اولین علت مرگ ناشی از سرطان را به خود اختصاص می‌دهد (۳). بر اساس برآورد انجمن سرطان امریکا در سال ۲۰۱۲ تعداد موارد جدید ابتلا به سرطان پستان در بین زنان در این کشور ۲۳۲،۳۴۰ نفر خواهد بود و تعداد مرگ در اثر سرطان پستان در زنان این کشور را ۳۹،۶۲۰ مورد برآورد کرده است (۴). به طور کلی یک نفر از هر هشت زن در طول عمر خود به سرطان پستان مبتلا می‌شود و از هر ۳۶ نفر یک نفر به علت سرطان پستان می‌میرد (۵).

روش درمان سرطان پستان ابتدا جراحی و سپس شیمی‌درمانی و در برخی موارد همراه رادیوتراپی است. در مواردی که بزرگی زیاد توده سرطانی ممکن است موجب برداشتن کل پستان بشود پیش از جراحی اقدام به شیمی‌درمانی می‌شود تا سایز تومور کوچک گردد که به آن نئوادجوانت تراپی گفته می‌شود. در برخی از سرطان‌های پستان زنان از هورمون تراپی و/یا آنتی بادیهای منوکلونال نیز برای درمان استفاده می‌شود.

رژیم‌های متعدد شیمی‌درمانی برای درمان بیماران مبتلا به سرطان پستان مورد بررسیهای دقیق کارآزمایی بالینی قرار گرفته است و ترکیبات مختلف این داروها برای بیماران مختلف مورد بررسی قرار گرفته‌اند. نتایج این پژوهش‌ها نشان می‌دهد که انتخاب نوع رژیم شیمی‌درمانی تصمیم مهمی است که بر درمان بیمار و میزان بقای او تاثیر می‌گذارد.

۱- توصیه‌های رژیم‌های درمانی

چهار رژیم درمانی برای درمان سرطان پستان وجود دارد که عموماً رژیم ۲ و ۴ در کشور ایران استفاده می‌گردد (6).

۱- یک رژیم همراه با کنترل چربی و کالری اضافی ممکن است مفید باشد. بر اساس برخی فرضیه‌ها، چاقی ممکن است رشد تومور ابتدایی را توسعه دهد و چون به نظر می‌رسد کالری، نقش مهم‌تری بازی می‌کند، لذا ۲۰٪ دریافت چربی، همراه با کاهش اسیدهای چرب اشباع شده (SFAs) تا ۷٪ را پیشنهاد می‌کنند.

۲- برای دسترسی به مواد مغذی مهم و فیتوکمیکال‌ها، مصرف روزانه حداقل ۹-۵ سهم میوه و سبزی و ۶ سهم غلات را باید تشویق کنید. هر سهم شامل یک قطعه میوه تازه، ۱۸۰ سی‌سی (۳/۴ فنجان) آب میوه ۱۰۰٪ طبیعی، ۱/۲ فنجان سبزی‌های پخته شده یا میوه‌های کنسرو شده، ۱ فنجان سبزی‌های سبز برگ یا سالاد، یک مشت (۱/۴ فنجان) میوه خشک شده و ۱/۲ فنجان نخود یا لوبیای خشک شده است.

۳- منابع رژیم غذایی ویتامین‌ها و مواد معدنی که مقدار دریافت رژیمی مرجع (DRI) و RDA را تأمین می‌کنند، معمولاً کافی هستند اما ممکن است مکمل معمولی نیز با اطمینان توصیه شود. خصوصیات شخصیتی و دموگرافی، زمان سپری شده پس از تشخیص و مرحله سرطان در هنگام تشخیص، متضمن پیشگویی مصرف مکمل رژیمی برای زنان در معرض خطر عود سرطان پستان است؛ مصرف مکمل در ۸۰٪ زنان دارای تاریخچه سرطان پستان، رایج است.

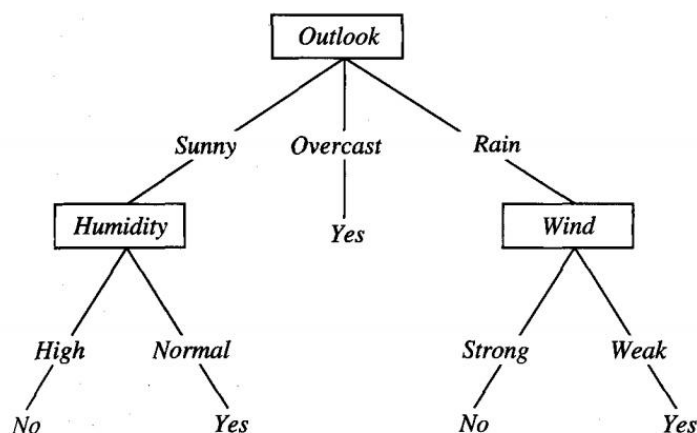
۴- دریافت سوپا به نظر می‌رسد برای زنان، عامل محافظتی باشد به جز آنهایی که به استروژن حساس هستند. دریافت روزانه سوپا به نظر نمی‌رسد یک عامل خطر برای پیشرفت سرطان باشد؛ مطالعات بیشتری در دست اقدام است.

۲- درخت تصمیم

درخت تصمیم یکی از پرکاربردترین و سودمندترین روش‌هایی است که برای استنتاج استقرایی مورد استفاده قرار می‌گیرد (۱۰). درخت تصمیم، یک روش غیر پارامتریک با ساختار سلسله مراتبی داده‌ها و یادگیری بانظارت است که با استفاده از استراتژی تقسیم و حل پیاده‌سازی می‌گردد. در این روش تابع یاد گرفته شده به صورت یک درخت تصمیم نمایش داده می‌شود که گاهی اوقات برای افزایش درجه خوانایی آن برای انسان، درخت را به صورت مجموعه‌ای از قوانین اگر-آنگاه در می‌آورند. این الگوریتم‌ها بر اساس داده‌های مجموعه یادگیری، یک درخت ایجاد می‌کنند در این درخت هر گره داخلی یک آزمون بر روی یک صفت را نشان می‌دهد، هر شاخه نتیجه‌ی ارزیابی را نشان می‌دهد و هر برگ بر چسب یک کلاس را نگهداری می‌کند. نمونه‌ای از درخت تصمیم گیری که از روی جدول ۱ ساخته شده است به همراه مجموعه قوانین معادل آن در شکل ۱ مشاهده می‌شود شکل ۲ ساختار یک درخت تصمیم را نشان می‌دهد. یک درخت تصمیم متشکل از یک سری گره‌های تصمیم داخلی و برگ‌های پایانی است.

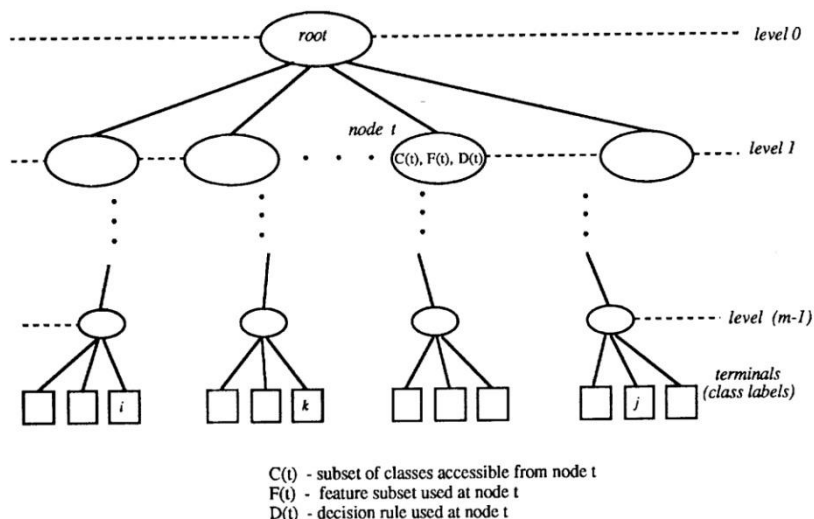
جدول ۱ داده‌های یادگیری مربوط به بازی تنیس (۱۰)

Day	Outlook	Temperature	Humidity	Wind	playTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	String	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Mild	High	Strong	Yes
D14	Rain	Mild	High	Strong	No



شکل ۱ نمونه‌ای از یک درخت تصمیم

هر گره تصمیم مانند t ، یک تابع آزمون $fm(t)$ را پیاده سازی می‌کند که شاخه‌ها را با خروجی‌های مجزا برچسب می‌زند. با ورودی مشخص، آزمون تست برای گره انجام می‌شود و یکی از شاخه‌ها با توجه به خروجی‌اش انتخاب می‌گردد. این رویه از ریشه شروع می‌شود و به صورت بازگشتی تا زمانی که به گره برگ (برگ‌های پایانی) برسد تکرار می‌گردد. در این مرحله، برچسب خروجی به این ورودی مشخص نسبت داده می‌شود.



شکل ۲ نمایشی از یک درخت تصمیم

هر تابع f ، فضای ورودی D بعدی را به نواحی کوچک‌تر تقسیم می‌کند. $f_m(\cdot)$ یک تابع است و زمانی که در یک درخت تصمیم بکار می‌رود تابعی پیچیده است که به چندین دنباله از تصمیم‌های ساده شکسته می‌شود. الگوریتم‌های مختلف درخت‌های تصمیم، مدل‌های متفاوتی برای $f_m(\cdot)$ در نظر می‌گیرند و هر مدل یک شکل متفاوت برای جدا کردن داده‌های هر گره به نواحی مختلف دارد. هر گره برگ دارای یک مقدار خروجی است که برای مسئله طبقه‌بندی، برچسب کلاس و برای مسئله رگرسیون، مقدار عددی در نظر گرفته می‌شود. هر گره برگ یک ناحیه محلی را با توجه به فضای ورودی تعریف می‌کند که نمونه‌های با برچسب‌های یکسان در آن ناحیه قرار می‌گیرند. در ادامه درخت‌های تک متغیر که فرآیند آزمون فقط روی یک متغیر ورودی صورت می‌گیرد معرفی می‌گردند و خواهیم دید که این درخت‌ها چگونه می‌توانند برای طبقه‌بندی مورد استفاده قرار گیرند.

۳- الگوریتم پیشنهادی

در این پروژه از داده‌های ۸۰۹ بیمار مبتلا به سرطان پستان و ۲۰ ویژگی از هر بیمار، استفاده شد. با توجه به گمشدگی نسبتاً زیاد داده‌ها در این مجموعه، تنها اطلاعات مربوط به ۶۸۳ بیمار قابل استفاده بود. بنابراین، در پایان‌نامه از ۶۸۳ رکورد مربوط به افرادی که مبتلا به سرطان پستان بودند استفاده شد. این ۲۰ ویژگی در جدول ۲ مشاهده می‌شود.

جدول ۲ ویژگی بیماران سرطان پستان

اطلاعات تفصیلی	نام	ردیف
سن بیماران بین ۴۰ تا ۸۰ سال است	Age	۱
وضعیت ازدواج • مجرد، • متأهل • مطلقه • بیوه	status Marital	۲
وضعیت یائسگی • قبل از یائسگی • بعد از یائسگی • اوایل یائسگی	mence status	۳
وضعیت تحصیلی ۱. بی‌سواد	mizan tahsilat	۴

۲. ابتدایی ۳. سیکل ۴. دیپلم ۵. دانشگاهی		
وضعیت جراحی • حفظ پستان • برداشت پستان • برداشتن مقداری از بافت	Surgery	۵
میزان شیمی درمانی: اعدادی بین ۱ تا ۹ است	Chemotherapy	۶
میزان پرتو درمانی: اعدادی بین ۱ تا ۹ است	Radio therapy	۷
دوز هورمون درمانی: اعداد بین ۱ تا ۹ است	Hormone therapy	۸
نوع هورمون درمانی: ۴ نوع دارو وجود دارد که با تشخیص دکتر استفاده شده است	Hormone therapy kind	۹
اندازه تومور به سانتی متر، بین ۱ تا ۱۰ سانت است.	Size of tumor	۱۰
اندازه تومور در غده لنفاوی	lymphatic tumor	۱۱
عود سرطان بعد از عمل	Relapse	۱۲
مرحله بالینی: بین ۱ تا ۵ است	clinical step	۱۳
مرحله بالینی جدید: ۵ روز بر مبنای روش‌های استاندارد وجود دارد ولی این ۵ روز دارای تغییراتی نیز است که در صورت تغییر ذکر شده است.	new clinical step	۱۴
HER-9 P53 PR ER	جواب چهار آزمایش	۱۹-۱۵
چهار رژیم درمانی وجود دارد که در قسمت ۲ مشاهده شده است که فقط رژیم درمانی ۲ و ۴ در ایران توصیه می‌شود	رژیم درمانی	۲۰

۳-۱ پیش‌پردازش

سپس سه مرحله پیش‌پردازش بر روی داده‌ها اعمال شده است که شامل مراحل زیر است.

۱- تخمین داده‌های گمشده

۲- کاهش ابعاد

۳- گسسته‌سازی

جهت اعمال پیش‌پردازش از نرم‌افزار SPSS.V20 استفاده گردید.

۳-۱-۱ تخمین داده‌های گمشده

جهت تخمین داده‌های گمشده از روش ماکزیمم درست‌نمایی^۱ (EM) استفاده شده است، معروف‌ترین روش آماری برای محاسبات داده‌های گمشده روش EM است، الگوریتم EM اولین بار در سال ۱۹۷۷ توسط دمپستر، لارد و روبین ارائه گردید، این الگوریتم بعدها در سال ۲۰۰۷ توسط مک لچلند توسعه یافت (۸)، EM یکی از معروف‌ترین الگوریتم‌های پیش‌بینی داده‌های گمشده است که در آن برای انتساب مقدار یک متغیر از سایر متغیرها استفاده می‌شود (Expectation) سپس الگوریتم می‌آزماید که آیا این مقدار محتمل‌ترین مقدار است (Maximization). اگر نباشد، یک مقدار محتمل‌تر انتساب می‌دهد. این کار آن قدر ادامه پیدا می‌کند تا به محتمل‌ترین مقدار برسد. EM تکنیکی مناسبی است که اغلب در تحلیل داده‌ها برای مدیریت داده‌های گمشده به کار می‌رود. در واقع EM قادر است بر برخی از محدودیت‌های سایر روش‌ها مانند میانگین و رگرسیون غلبه کند. برخلاف روش‌های میانگین و رگرسیون، روش EM خطای استاندارد^۲

^۱ Expectation Maximization

^۲ -Standard errors

خطای استاندارد، انحراف معیار یک توزیع نمونه‌برداری آماری است که برای تخمین انحراف معیار به دست آمده از تعدادی نمونه کاربرد دارد(۹) را در مسئله لحاظ می‌کند(۱۰).

۳-۱-۲ کاهش ابعاد

جهت کاهش ابعاد از یکی از معروف‌ترین الگوریتم‌های بنام PCA استفاده شده است و تعداد ابعاد به ۹ بعد کاهش یافت. تکنیک PCA بهترین روش برای کاهش ابعاد داده به صورت خطی است. یعنی با حذف ضرایب کم اهمیت به دست آمده از این تبدیل، اطلاعات از دست رفته نسبت به روش‌های دیگر کمتر است. البته کاربرد PCA محدود به کاهش ابعاد داده نمی‌شود و در زمینه‌های دیگری مانند شناسایی الگو و تشخیص چهره نیز مورد استفاده قرار می‌گیرد. در این روش محورهای مختصات جدیدی برای داده‌ها تعریف شده و داده‌ها بر اساس این محورهای مختصات جدید بیان می‌شوند. اولین محور باید در جهتی قرار گیرد که واریانس داده‌ها ماکسیمم شود (یعنی در جهتی که پراکندگی داده‌ها بیشتر است). دومین محور باید عمود بر محور اول به گونه‌ای قرار گیرد که واریانس داده‌ها ماکسیمم شود. به همین ترتیب محورهای بعدی عمود بر تمامی محورهای قبلی به گونه‌ای قرار می‌گیرند که داده‌ها در آن جهت دارای بیشترین پراکندگی باشند.

۳-۱-۴ گسسته‌سازی

داده‌ها پس تخمین و کاهش ابعاد تبدیل به اعداد پیوسته می‌شود ولی الگوریتم درخت تصمیم نیاز به داده‌های گسسته دارد. لذا از نرم افزار SPSS برای گسسته‌سازی اطلاعات استفاده گردید.

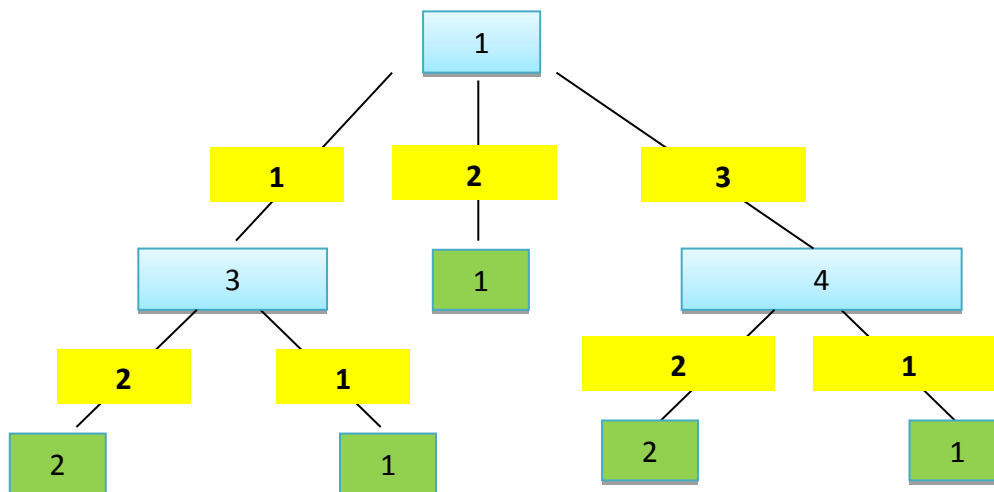
۲-۴ حل مسئله به کمک درخت تصمیم

برای حل مسئله به روش درخت تصمیم ابتدا یک ساختمان داده مناسب طراحی می‌کنیم که به صورت زیر می‌باشد.

۱-۲-۴ طراحی ساختمان داده برای درخت تصمیم

جهت ذخیره درخت تصمیم از یک آرایه دوبعدی استفاده می‌گردد که تعداد ستون‌های آن به اندازه گره‌های درخت تصمیم است و تعداد سطرهای آن سه می‌باشد. به عنوان مثال برای ذخیره سازی درخت تصمیم شکل ۱ نیاز به آرایه‌ای با ۸ ستون و ۳ سطر است. سطر اول آرایه در دو حالت ذخیره می‌گردد، اگر مقدار آن مثبت باشد به معنی گره‌های غیر برگ است و اگر مقدار آن منفی باشد به معنی گره برگ است. به عنوان مثال در شکل ۳ ستون‌های ۱، ۲ و ۶ نشان دهنده گره‌های غیر برگ است و باقی ستون‌های نشان دهنده گره‌های برگ است.

سطر دوم اشاره به اندیس والد هر گره است، به عنوان مثال در شکل ۳ در ستون دوم، مقدار سطر دوم آن برابر یک است و نشان می‌دهد والد گره سوم، در ستون اول قرار دارد. و اگر ستون دوم دارای مقدار صفر باشد به معنی گره ریشه است. سطر سوم نشان دهند مقدار یال درخت است، به عنوان مثال در شکل ۳ در ستون ششم مقدار سطر سوم آن برابر سه است و نشان می‌دهد که یال این گره دارای مقدار سه است. و اگر ستون سوم دارای مقدار صفر باشد به معنی گره ریشه است.



	۱	۲	۳	۴	۵	۶	۷	۸
۱	۱	۳	-۲	-۱	-۱	۴	-۲	-۱
۲	۰	۱	۲	۲	۱	۱	۶	۶
۳	۰	۱	۲	۱	۲	۳	۲	۱

شکل ۳ ساختار پیشنهادی برای ذخیره سازی درخت تصمیم

نحوه ساخت درخت تصمیم Inorder می‌باشد، بدین معنی ابتدا سمت ریشه ساخته می‌شود سپس گره سمت چپ و در آخر گره سمت راست ساخته می‌شود. در درخت تصمیم ملاک انتخاب Entropy است. شرط پایان الگوریتم درخت تصمیم یکی از دو شرط زیر است:

دیگر نمونه‌ای وجود نداشته باشد.

دیگر ویژگی نداشته باشیم (استفاده از قانون رای گیری اکثریت)

۴- ارزیابی

یکی از مهم‌ترین قسمت‌های یک نظریه انجام آزمایشات و اثبات نتایج آن‌ها است. ما برای تست نظریه ارائه داده شده برنامه‌هایی با زبان متلب ایجاد نمودیم که در ادامه به توضیح و نمایش نتایج بدست آمده از آن‌ها می‌پردازیم. این آزمایشات روی سیستمی با پردازنده 4GH و حافظه 6GB انجام شده است، برای تست این داده‌ها از اطلاعات یک سال دانشگاه سما ماهشهر، استفاده شده است.

۴-۱ تقسیم داده‌های به آموزشی و تست

داده‌های الگوریتم را به سه روش به داده‌های آموزشی و تست تقسیم می‌کنیم

• شست درصد داده‌ها جهت آموزشی و چهل درصد داده‌ها جهت تست

• هفتاد درصد داده‌ها جهت آموزش و سی درصد داده‌ها جهت تست

• استفاده از روش K-Fold و مقادیر K اعدادی بین ۲ الی ۱۰ در نظر گرفته می‌شود.

سپس الگوریتم پیشنهادی اجرا می‌کنیم و چهار پارامتر Matrix Confusion، Sensitivity، Specificity و Accuracy مورد بررسی قرار داده شد، سپس زمان اجرا هر کدام نیز بررسی می‌گردد. و خروجی آن در جدول ۳ مشاهده می‌شود.

همانگونه که در جدول ۳ مشاهده می‌شود بهترین دقت در روش درخت تصمیم در روش 2-Fold با دقت ۹۲،۶۷ درصد رسیده شد.

جدول ۳ مقایسه الگوریتم پیشنهادی از چهار معیار و زمان اجرا برای داده‌های آموزش و تست

ردیف	تقسیم بندی داده آموزشی و تست	درخت تصمیم				زمان بر حسب ثانیه
		Accuracy	Specificity	Sensitivity	confusion Matrix	
۱	آموزشی	%۱۰۰	%۱۰۰	%۱۰۰	۲۶۷	۰
					۰	۱۴۲
	تست	%۸۹،۰۱	%۸۹،۲۵	%۸۸،۵۱	۱۶۶	۱۰
					۲۰	۷۷
۲	آموزشی	%۱۰۰	%۱۰۰	%۱۰۰	۳۰۸	۰
					۰	۱۷۰
	تست	%۹۰،۶۹	%۸۶،۷۶	%۹۲،۶۵	۱۲۶	۹
					۱۰	۵۹
۳	2-Fold	%۱۰۰	%۱۰۰	%۱۰۰	۲۲۵	۰
					۰	۱۱۷
	تست	%۹۲،۶۷	%۸۸،۸۰	%۹۴،۹۱	۲۰۵	۱۴

				۱۱	۱۱۱			
۰,۱۸۹۲۶۹	%۱۰۰	%۱۰۰	%۱۰۰	۲۴۵	۰	آموزشی	3-Fold	۴
				۰	۱۳۵			
				۱۸۷	۱۲			
۰,۲۸۲۶۸۶	%۱۰۰	%۱۰۰	%۱۰۰	۱۵	۸۹	تست	4-Fold	۵
				۲۷۶	۰			
				۰	۱۵۱			
۰,۳۷۳۳۲۳	%۱۰۰	%۱۰۰	%۱۰۰	۱۵۹	۱۰	آموزشی	5-Fold	۶
				۰	۱۶۵			
				۱۳۸	۷			
۰,۵۱۵۲۵۵	%۱۰۰	%۱۰۰	%۱۰۰	۱۴	۷۴	تست	6-Fold	۷
				۳۰۰	۰			
				۰	۱۶۵			
۰,۵۶۴۴۷۸	%۱۰۰	%۱۰۰	%۱۰۰	۱۳۸	۷	آموزشی	7-Fold	۸
				۰	۱۷۴			
				۱۱۹	۵			
۰,۶۴۹۵۹۹	%۱۰۰	%۱۰۰	%۱۰۰	۱۱	۵۴	تست	8-Fold	۹
				۳۳۴	۰			
				۰	۱۸۲			
۰,۷۳۱۵۱۳	%۱۰۰	%۱۰۰	%۱۰۰	۱۰۶	۴	آموزشی	9-Fold	۱۰
				۰	۱۹۳			
				۸۰	۹			
۰,۸۳۹۹۳۰	%۱۰۰	%۱۰۰	%۱۰۰	۹	۴۱	تست	10-Fold	۱۱
				۳۴۶	۰			
				۰	۱۸۸			
۰,۶۴۹۵۹۹	%۱۰۰	%۱۰۰	%۱۰۰	۹۴	۵	آموزشی	8-Fold	۹
				۰	۱۸۸			
				۱۰	۴۱			
۰,۷۳۱۵۱۳	%۱۰۰	%۱۰۰	%۱۰۰	۳۵۵	۰	آموزشی	9-Fold	۱۰
				۰	۱۹۳			
				۸۰	۹			
۰,۸۳۹۹۳۰	%۱۰۰	%۱۰۰	%۱۰۰	۹	۳۷	تست	10-Fold	۱۱
				۳۶۳	۰			
				۰	۱۹۷			
۰,۸۳۹۹۳۰	%۱۰۰	%۱۰۰	%۱۰۰	۷۲	۱۰	آموزشی	10-Fold	۱۱
				۰	۱۹۷			
				۷	۳۵			

جهت ترکیب داده‌های آموزشی و تست به یک داده، از روابط زیر استفاده می‌شود. این رابطه به این دلیل انتخاب شده که اهمیت داده تست نسبت به آموزشی بیشتر است.

$$Sensitivity = \frac{Sensitivity_{Training} + 3 \times Sensitivity_{Test}}{4} \quad (1)$$

$$Specificity = \frac{Specificity_{Training} + 3 \times Specificity_{Test}}{4} \quad (2)$$

$$Accuracy = \frac{Accuracy_{Training} + 3 \times Accuracy_{Test}}{4} \quad (3)$$

جدول ۴ ترکیب داده‌های آموزشی و تست برای الگوریتم را نشان می‌دهد. و بهترین دقت اجرای الگوریتم با رنگ دیگر تمایز شده است.

جدول ۴ مقایسه الگوریتم پیشنهادی از برای حسب سه معیار

درخت تصمیم			تقسیم بندی داده آموزشی و تست	ردیف
Accuracy	Specificity	Sensitivity		
%۹۱,۷۵۷۵	%۹۱,۹۳۷۵	%۹۱,۳۸۲۵	۴۰-۶۰	۱
%۹۴,۴۸۷۵	%۸۷,۰۷۰۰	%۹۳,۰۱۷۵	۳۰-۷۰	۲
%۹۴,۵۰۲۵	%۹۱,۶	%۹۶,۱۸۲۵	2-Fold	۳
%۹۳,۳۲	%۹۱,۰۹	%۹۴,۴۲۷۵	3-Fold	۴
%۹۳	%۹۱,۰۸	%۹۳,۹۳	4-Fold	۵
%۹۳,۱۵	%۹۲,۲۸	%۹۳,۵۴	5-Fold	۶
%۹۳,۶۵	%۹۳,۶۵	%۹۳,۶۶	6-Fold	۷
%۹۳,۷۳	%۹۴,۱۲	%۹۳,۵۴	7-Fold	۸
%۹۲,۵	%۹۱,۸۵	%۹۲,۷۹	8-Fold	۹
%۹۰,۰۷	%۸۵,۳۲	%۹۲,۴۲	9-Fold	۱۰
%۸۹,۷۲	%۸۳,۳۴	%۹۳,۳۶	10-Fold	۱۱

نتیجه گیری

در این پژوهش نشان داده شد که جهت دسته بندی کردن اطلاعات بیماری نیاز به دسته بندی کننده غیر خطی می‌باشد، و مشخص شد جهت دسته بندی غیر خطی می توان از و درخت تصمیم استفاده کرد. همچنین مشخص شد جهت تشخیص نوع تغذیه درمانی برای بیماران سرطان پستان ارائه شد که نشان دادیم می‌تواند درصد بالاتری را برای ما فراهم می‌نماید.

منابع و مراجع

- [1] Cancer - United States Cancer Statistics (USCS) Data - 2009 Top Ten Cancers [Internet]. [cited 2013 Sep 13]. Available from: <http://apps.nccd.cdc.gov/uscs/toptencancers.aspx#text>
- [2] Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, et al. Cancer statistics, 2006. *CA Cancer J Clin*. 2006;56(2):106-30.
- [3] GHE_DthWHOReg6_2000_2011.xls [Internet]. [cited 2013 Sep 13]. Available from: http://www.who.int/entity/healthinfo/global_burden_disease/GHE_DthWHOReg6_2000_2011.xls
- [4] CDC - Cancer Among Women [Internet]. [cited 2013 Sep 13]. Available from: <http://www.cdc.gov/cancer/dcpc/data/women.htm>
- [5] How many women get breast cancer? [Internet]. [cited 2013 Sep 13]. Available from: <http://www.cancer.org/cancer/breastcancer/overviewguide/breast-cancer-overview-key-statistics>
- [6] Mousavi SM, Montazeri A, Mohagheghi MA, Jarrahi AM, Harirchi I, Najafi M, et al. Breast cancer in Iran: an epidemiological review. *Breast J*. 2007;13(4):383-91.
- [7] Harirchi I, Kolahdoozan S, Karbakhsh M, Chegini N, Mohseni S, Montazeri A, et al. Twenty years of breast cancer in Iran: downstaging without a formal screening program. *Ann Oncol*. 2011;22(1):93-7.
- [8] Stockler MR, Harvey VJ, Francis PA, Byrne MJ, Ackland SP, Fitzharris B, et al. Capecitabine Versus Classical Cyclophosphamide, Methotrexate, and Fluorouracil As First-Line Chemotherapy for Advanced Breast Cancer. *J Clin Oncol*. 2011 Dec 1;29(34):4498-504.
- [9] Jones SE, Erban J, Overmoyer B, Budd GT, Hutchins L, Lower E, et al. Randomized Phase III Study of Docetaxel Compared With Paclitaxel in Metastatic Breast Cancer. *J Clin Oncol*. 2005 Aug 20;23(24):5542-51.
- [10] Martín M, Ruiz A, Borrego MR, Barnadas A, González S, Calvo L, et al. Fluorouracil, Doxorubicin, and Cyclophosphamide (FAC) Versus FAC Followed by Weekly Paclitaxel As Adjuvant Therapy for High-Risk, Node-Negative Breast Cancer: Results From the GEICAM/2003-02 Study. *J Clin Oncol*. 2013 Jul 10;31(20):2593-9.
- [11] Jones SE, Savin MA, Holmes FA, O'Shaughnessy JA, Blum JL, Vukelja S, et al. Phase III Trial Comparing Doxorubicin Plus Cyclophosphamide With Docetaxel Plus Cyclophosphamide As Adjuvant Therapy for Operable Breast Cancer. *J Clin Oncol*. 2006 Dec 1;24(34):5381-7.
- [12] Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879-86.
- [13] Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*. 1994;73(3):643-51.
- [14] Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004;23(7):1111-30.
- [15] Lavra\vc N. Selected techniques for data mining in medicine. *Artif Intell Med*. 1999;16(1):3-23