

کاربرد تحلیل منحنی مشخصه عملکرد تعدیل یافته در ترکیب نشانگرهای زیستی برای تشخیص زودهنگام دیابت حاملگی

مأنده امینی^۱ (Ph.D Candidate)، انوشیروان کاظم‌نژاد^{۱*} (Ph.D)، فرید زابری^۲ (Ph.D)، اعظم امیریان^۳ (M.Sc)، نورالسادات کریمان^۴ (Ph.D)

۱- گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، تهران، ایران

۲- مرکز تحقیقات پروتئومیکس، گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۳- گروه مامایی، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی جیرفت، جیرفت، ایران

۴- گروه مامایی و بهداشت باروری، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

تاریخ دریافت: ۱۳۹۷/۸/۵ تاریخ پذیرش: ۱۳۹۸/۲/۱۰

kazem_an@modares.ac.ir

* نویسنده مسئول، تلفن: ۰۲۱-۸۲۸۸۳۸۷۵

چکیده

هدف: در حوزه پزشکی تشخیصی، ارزیابی دقت تشخیصی نشانگرهای زیستی (بیومارکرها) یا تست‌ها همواره مورد توجه بوده است. گاهی ممکن است یک نشانگر زیستی به تنهایی برای پیش‌بینی بیماری به اندازه کافی حساس و ویژه نباشد. اما با ترکیب چندین نشانگر شاید بتوان به تشخیص بهتری دست یافت. هدف از این مطالعه، بررسی عملکرد ترکیب نشانگرهای زیستی برای تشخیص زودهنگام دیابت بارداری بود.

مواد و روش‌ها: در این مقاله، از اطلاعات ۵۲۳ زن باردار مراجعه‌کننده به بیمارستان‌های مهدیه و طالقانی واقع در شهر تهران استفاده شد. میزان استریول غیر کونژوگه، آلفا فیتوپروتئین و بتا گنادوتروپین کوریونی انسانی در اوایل سه ماهه دوم بارداری مورد اندازه‌گیری قرار گرفت. از سطح زیر منحنی مشخصه عملکرد و معیارهای حساسیت و ویژگی برای ارزیابی دقت این نشانگرها و نیز یافتن ترکیب خطی بهینه‌ای از آن‌ها استفاده شد. انتخاب نقطه برش بهینه بر اساس شاخص یودن بود.

یافته‌ها: در ترکیب دوتایی نشانگرها، بزرگ‌ترین مقدار سطح زیر منحنی مشخصه عملکرد، حساسیت و ویژگی به ترتیب ۵۹۳/۰، ۶۱/۹۰ درصد و ۵۸/۳۰ درصد با نقطه برش بهینه ۰/۱۱ به دست آمد. در ترکیب سه‌تایی با تعدیل اثر سن و شاخص توده بدنی، بیش‌ترین مقدار سطح زیر منحنی مشخصه عملکرد، حساسیت و ویژگی به ترتیب ۰/۷۵۱، ۸۲/۹۵ درصد و ۷۴/۶۲ درصد با نقطه برش بهینه ۰/۱۰ محاسبه شد.

نتیجه‌گیری: بر اساس نتایج این پژوهش، ترکیب خطی سه‌تایی نشانگرها با در نظر گرفتن اثر متغیرهای کمکی به طور معناداری عملکرد تشخیصی بهتری داشت.

واژه‌های کلیدی: نشانگر زیستی، منحنی مشخصه عملکرد، دیابت بارداری، تشخیص اولیه

مقدمه

می‌شود [۵]. این منحنی یک روش گرافیکی بسیار پرکاربرد در حوزه‌هایی مانند رادیولوژی، بیوانفورماتیک، رادیولوژی، ماشین یادگیری و داده کاوی است [۶].

در عمل، بسیار اتفاق می‌افتد که چندین نشانگر زیستی را به‌طور هم‌زمان برای هر فرد اندازه‌گیری کنند. در چنین شرایطی، سؤال مهم آن است که آیا یک نشانگر به تنهایی و یا در ترکیب با سایر نشانگرها می‌تواند پیشگوی خوبی برای تشخیص زودرس بیماری مورد نظر باشد یا خیر؟ با مروری بر مطالعات گذشته روشن است که برخی محققان بر این باورند که با ترکیب نشانگرهای زیستی به صورت هم‌زمان و یافتن یک ترکیب بهینه

در آزمایش‌های تشخیصی پزشکی، ارزیابی دقت تست‌های آزمایشگاهی و نشانگرهای زیستی در تشخیص یک بیماری اهمیت حیاتی دارد. در واقع این موضوع می‌تواند به تصمیم‌گیری‌های بهتر برای جلوگیری از عوارض احتمالی و استفاده از درمان‌های مناسب کمک کند [۱]. طی سال‌های متعددی از روش‌های مختلف آماری برای ارزیابی دقت تست‌های تشخیصی به صورت جداگانه استفاده شده است [۲-۴]. زمانی که نتیجه یک تست تشخیصی در مقیاس رتبه‌ای و یا پیوسته باشد، استفاده از منحنی راک برای ارزیابی دقت تشخیص پیشنهاد

یک منهای ویژگی (نرخ مثبت کاذب) محاسبه شده و سپس نقاط به دست آمده به یک دیگر وصل می‌شوند. سطح زیر منحنی راک (AUC) یک شاخص تک مقدار بوده که از آن به طور گسترده برای ارزیابی قدرت تشخیص تست‌ها استفاده می‌شود. به طور کلی مساحت زیر منحنی راک بیانگر قدرت تست تشخیصی و یک معیار مؤثر و ترکیبی از حساسیت و ویژگی بوده که اعتبار ذاتی تست‌های تشخیصی را توصیف می‌کند. این شاخص از رابطه زیر محاسبه می‌شود:

$$AUC = \int_0^1 ROC(t) dt.$$

دامنه تغییرات AUC بین ۰ و ۱ است. مقدار بزرگ‌تر AUC بیانگر توانایی بیش‌تر نشانگرها در تمایز افراد سالم از بیمار برای همه مقادیر ممکن نقاط برش است. زمانی که تست تشخیصی کامل (استاندارد طلایی) باشد، منحنی راک در نقطه برش t برابر ۱ و در نتیجه $AUC=1$ خواهد شد. این بدان معنا است که تست مورد نظر در تمایز بین افراد بیمار و غیر بیمار کاملاً خوب عمل کرده است [۲۹].

برای برآورد دقت رده‌بندی با استفاده از روش‌های استاندارد منحنی راک، وضعیت بیماری برای هر فرد بدون خطا اندازه‌گیری می‌شود. وضعیت واقعی بیماری را استاندارد طلایی می‌نامند. استاندارد طلایی اغلب می‌تواند با پیگیری در طول مطالعه، کالبد شکافی و غیره که مورد تأیید کمیته کارشناسان قرار گرفته در دسترس باشد. در این حالت، برای هر فرد یک جدول توافقی دو در دو در نظر گرفته می‌شود که شامل چهار ترکیب از رده‌بندی و وضعیت واقعی بیماری است. در سلول‌های این جدول تعداد منفی‌های واقعی، منفی‌های کاذب، مثبت‌های واقعی و مثبت‌های کاذب قرار دارد. استفاده از تحلیل منحنی راک دارای چند مزیت است: اول این‌که در مقایسه با معیارهای حساسیت و ویژگی، سطح زیر منحنی راک تحت تأثیر معیار تصمیم‌گیری نبوده و مستقل از شیوع بیماری است. دوم این‌که چندین تست تشخیصی می‌توانند به صورت هم‌زمان با استفاده از سطح زیر منحنی راک با یک دیگر مقایسه شوند. سوم این‌که نقطه برش بهینه را می‌توان در تحلیل منحنی راک تعیین نمود. برای یافتن نقطه برش بهینه می‌توان شاخص یودن را به کار برد. این شاخص مجموع نرخ رده‌بندی صحیح را برای دو رده سالم و بیمار بیشینه می‌کند. بر پایه نتایج برخی مطالعات، با ترکیب بیومارکرها و استفاده از شاخص یودن، نمره ترکیبی حاصل دارای بزرگ‌ترین نرخ کلی رده‌بندی صحیح در نقطه برش انتخابی در بین همه ترکیبات خطی ممکن خواهد بود [۳۰].

اغلب ممکن است تشخیص و پیش‌بینی بیماری‌ها با یک نشانگر زیستی عملکرد خوبی نداشته باشد. لذا در این نوع مطالعات در صورت امکان، تشخیص با بیش از یک نشانگر انجام

قدرت تمایز بین افراد سالم و بیمار افزایش می‌یابد [۷-۱۲]. روش‌های متعددی برای ترکیب از جمله تحلیل جداسازی، رگرسیون لجستیک و رگرسیون منطقی وجود دارد [۹، ۱۳، ۱۴]. اما شاید یکی از ساده‌ترین روش‌ها، ترکیب نشانگرها به صورت خطی باشد. این روش با به‌کارگیری AUC به عنوان تابع هدف، به طور گسترده‌ای در مطالعات مختلف مورد بررسی قرار گرفته است [۱۵].

برای تشخیص و یا تخصیص هر فرد به طبقه سالم یا بیمار، نیاز به تعیین یک مقدار آستانه (نقطه برش) داریم. دقت تشخیصی معمولاً با استفاده از نرخ طبقه‌بندی صحیح کلی (یعنی مجموع حساسیت و ویژگی) در یک مقدار آستانه مشخص اندازه‌گیری می‌شود [۱۹]. AUC یک معیار خلاصه کلی از دقت تشخیصی در همه نقاط برش ممکن است. بنابراین برای یافتن یک ترکیب خطی بهینه مطلوب و دستیابی به بزرگ‌ترین نرخ کلی رده‌بندی صحیح، نرخ کلی رده‌بندی صحیح نیز باید مورد توجه قرار بگیرد. به طور کلی برای انتخاب نقطه برش در یک مقدار مشخص بیومارکرها، روش‌های مختلفی وجود دارد که از بین آن‌ها شاخص یودن محبوب‌ترین آن‌ها است [۲۰].

دیابت بارداری از عوارض شایع دوران بارداری بوده که می‌تواند بر مادر و جنین در دوران قبل و بعد از زایمان اثرات نامطلوب داشته باشد [۲۴]. از این رو تشخیص زودرس و انجام مداخله‌های مناسب شاید بتواند تا اندازه‌ای از بروز بیماری‌هایی مانند سندرم داون جلوگیری کند. در تعدادی مطالعه، رابطه بین نشانگرهای استریول غیر کونژوگه (Unconjugated estriol, uE3)، آلفا فیتوپروتئین (Alpha-Fetoprotein, AFP) و بتای گنادوتروپین کوریونی انسان (Beta-Human Chorionic Gonadotropin, β -HCG) با دیابت بارداری و سایر بیماری‌های مرتبط با حاملگی پرداخته شده است [۲۵-۲۸]. نتایج حاصل از این مطالعات حکایت از کم بودن دقت تشخیصی هر یک از این نشانگرها داشت. بنابراین هدف از انجام این پژوهش، بررسی قدرت ترکیب بهینه در پیش‌بینی زودرس دیابت بارداری از بین ترکیب‌های دوتایی و سه تایی این سه نشانگر زیستی با تعدیل اثر متغیرهای پیشگوی مؤثر در هفته ۱۴ تا ۱۷ حاملگی بود.

مواد و روش‌ها

منحنی راک برای ارزیابی توانایی تشخیصی نشانگرهای زیستی در تمایز وضعیت واقعی افراد، انتخاب یک نقطه برش بهینه و نیز مقایسه عملکرد تشخیصی استفاده می‌شود. برای رسم منحنی راک به صورت تجربی پس از اندازه‌گیری تست تشخیصی مورد نظر روی نمونه‌ای از افراد سالم و بیمار، نقاط برش مختلف تعیین و در هر نقطه برش مقادیر حساسیت (نرخ مثبت واقعی) و

می‌شود. در این صورت می‌توان اطلاعات تشخیصی به‌دست آمده از ترکیب چندین نشانگر را با اطلاعات حاصل از یک نشانگر با یک‌دیگر مقایسه کرد. هنگام استفاده از چندین نشانگر به‌صورت هم‌زمان، مسأله مهم چگونگی ترکیب اطلاعات مربوط به نتایج حاصل از آن‌ها برای تمایز بین افراد سالم از بیمار است. هم‌چنین برخی محققان با در نظر گرفتن این واقعیت که هزینه خطا در رده‌بندی‌های مختلف متفاوت است معتقدند که بهتر است هزینه کلی رده‌بندی موقع ترکیب نشانگرهای زیستی می‌نیم شود. اولین بار فیشر در سال ۱۹۳۶ موضوع جداسازی را با انتخاب ضرایب مطرح کرد که نسبت اختلاف میانگین ترکیب‌های خطی را در دو گروه به واریانس آن‌ها ماکزیم کند. طبق این نظریه، بهترین ترکیب خطی دوتایی زمانی منجر به ماکزیم شدن حساسیت می‌شود که توزیع هر دو نشانگر از نرمال پیروی کند و ماتریس‌های کوواریانس متناسب باشند [۱۶]. در ادامه به‌طور خلاصه به نحوه برآورد AUC در حالت ترکیب نشانگرهای زیستی اشاره می‌شود.

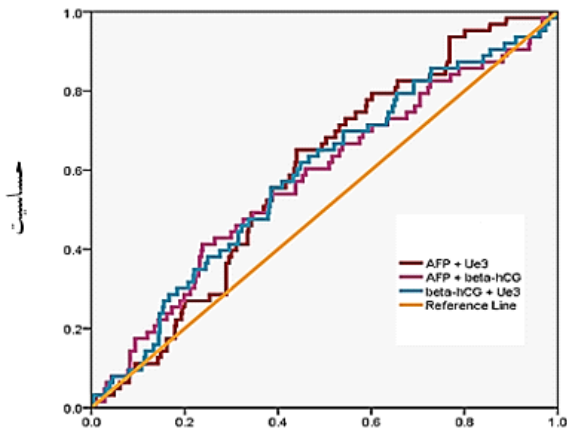
فرض کنید در مطالعه‌ای Y_1, Y_2, \dots, Y_k نتایج حاصل از k نشانگر زیستی و S نمره حاصل از این ترکیب باشد. در این صورت در ترکیب خطی $S(\alpha, Y) = \sum \alpha_p Y_p$ با انتخاب یکی از ضرایب $(\alpha_1, \dots, \alpha_p)$ می‌توان به ماکزیم دقت تشخیصی و سطح زیر منحنی بزرگ‌تری دست یافت. سو و لیو در سال ۱۹۹۳ مقدار AUC را برای ترکیب خطی همبسته $\alpha_1 Y_1 + \alpha_2 Y_2$ برآورد کردند. در روش آن‌ها، Y_1 و Y_2 دارای توزیع نرمال با پارامترهای میانگین و واریانس به ترتیب (μ^D, Σ^D) و $(\mu^{\bar{D}}, \Sigma^{\bar{D}})$ در دو رده سالم (D) و بیمار (\bar{D}) بودند. هاوونگ و همکاران نیز روشی را برای برآورد سطح زیر منحنی راک در زمانی که بیش از دو نشانگر زیستی در مطالعه وجود داشته باشد، در حالت ترکیب تست‌های تشخیصی ارائه دادند [۱۵].

برای برآورد منحنی راک و معیارهای تشخیصی از دو رویکرد پارامتری و ناپارامتری استفاده می‌شود [۳۱]. در صورت استفاده از روش‌های پارامتری، ابتدا باید نرمال بودن توزیع هر یک از نشانگرها و سپس نرمال چندمتغیره بودن آن‌ها را به‌هنگام ترکیب در هر یک از رده‌های سالم و بیمار بررسی نمود. در صورتی که با انجام تبدیلات، نرمال بودن توزیع حاصل نشد، می‌توان از روش‌های ناپارامتری و یا نیمه‌پارامتری استفاده کرد. داده‌های این پژوهش برگرفته از اطلاعات ۵۲۳ زن باردار در محدوده سنی ۲۰ تا ۴۰ سال بود که برای انجام تست‌های حاملگی به درمانگاه‌های مراقبت‌های دوران بارداری وابسته به دانشگاه علوم پزشکی شهید بهشتی تهران (بیمارستان‌های مهدیه و طالقانی) در سال ۱۳۹۶ مراجعه کرده بودند. میزان‌های $uE3$ و AFP بر اساس نانوگرم بر میلی‌لیتر و میزان β -HCG بر حسب

نتایج

در تحقیق حاضر از بین ۵۲۳ خانم باردار، ۶۳ نفر (۱۲ درصد) مبتلا به دیابت و ۴۶۰ نفر (۸۸ درصد) سالم بودند. میانگین \pm انحراف معیار ($uE3$) در دو گروه با و بدون دیابت بارداری به ترتیب $1/15 (\pm 0/60)$ و $1/05 (\pm 0/58)$ نانوگرم بر میلی‌لیتر بود. هم‌چنین میانگین \pm انحراف معیار (β -HCG) در مادران مبتلا به دیابت بارداری $1/09 (\pm 0/80)$ mIU/ml و در مادران غیر مبتلا $1/18 (\pm 0/76)$ mIU/ml بود. از طرفی میانگین \pm انحراف معیار (AFP) به تفکیک وضعیت ابتلا به دیابت بارداری به ترتیب $1/13 (\pm 0/36)$ و $1/11 (\pm 0/43)$ نانوگرم بر میلی‌لیتر گزارش شده است. میانگین \pm انحراف معیار سن مادران باردار نیز در دو گروه با و بدون دیابت به ترتیب $30/20 (\pm 4/78)$ و $28/56 (\pm 3/38)$ سال بود. میانگین \pm انحراف معیار شاخص توده بدنی در مادران باردار مبتلا و غیر مبتلا به دیابت به ترتیب $25/98 (\pm 5/57)$ و $24/38 (\pm 3/26)$ کیلوگرم بر متر مربع بود. با

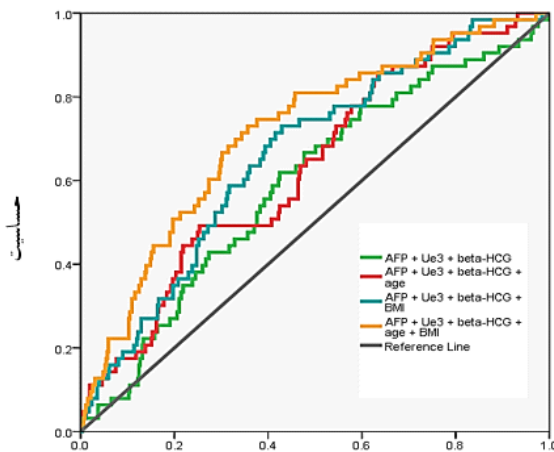
این‌رو در گام بعد برای بالا بردن دقت تشخیص ترکیب سه‌تایی نشانگرها مد نظر قرار گرفت.



ویژگی - ۱

شکل ۲. مقایسه سطح زیر منحنی راک برای ترکیب‌های دوتایی نشانگرهای زیستی در غربالگری دوم

طبق گزارش جدول ۳ نتایج نشان داد با تعدیل اثر پیشگوه‌های سن و شاخص توده بدنی در ترکیب سه‌تایی نشانگرها قدرت تشخیص به‌طور معناداری افزایش پیدا می‌کند ($AUC=0.751$ ، حساسیت برابر با $82/95$ درصد و ویژگی برابر با $74/62$ درصد). با ملاحظه شکل ۳ نیز می‌توان به‌طور شهودی دریافت که ترکیب همینه فوق برای پیش‌بینی زودرس دیابت بارداری در مقایسه با سایر ترکیبات مناسب‌تر است.

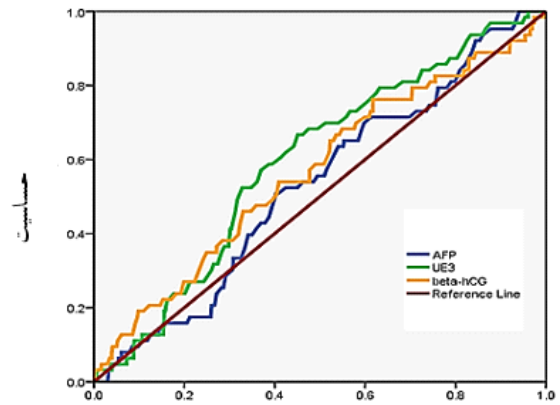


ویژگی - ۱

شکل ۳. مقایسه سطح زیر منحنی راک برای ترکیب‌های سه‌تایی نشانگرهای زیستی در غربالگری دوم

استفاده از آزمون من- ویتنی، بین دو گروه با و بدون دیابت بارداری تنها از نظر میانگین سن، شاخص توده بدنی و استریول غیر کونژوگه اختلاف معنادار وجود داشت ($P<0.05$).

با رسم منحنی راک هر یک از نشانگرهای β -uE3 و HCG و مقدار AUC به ترتیب برابر با 0.597 ، 0.567 و 0.531 به‌دست آمد (جدول ۱). هر چند استریول غیرکونژوگه در مقایسه با دو نشانگر دیگر مقدار AUC بزرگ‌تر و ارتباط معنادار با دیابت بارداری ($P=0.001$) داشت، اما با توجه به نزدیک بودن منحنی راک به قطر مربع از توان پیش‌بینی خوبی برخوردار نبود (شکل ۱). در مرحله بعد، ترکیب نشانگرها برای بالا بردن دقت تشخیص مورد بررسی قرار گرفت.



ویژگی - ۱

شکل ۱. مقایسه سطح زیر منحنی راک برای هر یک از نشانگرهای زیستی در غربالگری دوم

قبل از برآورد منحنی راک ترکیبی، ابتدا نرمال بودن توزیع هر یک از نشانگرها ارزیابی شد. نتایج نشان داد که هیچ یک از سه نشانگر دارای توزیع نرمال نبودند ($P<0.001$). با انجام تبدیل‌های لگاریتمی و باکس-کاکس روی هر سه نشانگر، توزیع آن‌ها نرمال شد ($P>0.05$). توزیع نرمال دومتغیره و چندمتغیره آن‌ها نیز تأیید شد ($P>0.05$). سپس بر اساس مقدار α محاسبه شده، مقدار ارزش‌های تشخیصی و نقطه برش همینه برای ترکیب‌هایی دوتایی نشانگرها به‌دست آمد (جدول ۲). همان‌طور که ملاحظه می‌شود هر چند ترکیب آلفا فیتوپروتئین و استریول غیر کونژوگه دارای بیش‌ترین مقدار AUC (0.593)، حساسیت ($61/90$ درصد) و ویژگی ($58/30$ درصد) بود، اما تست ترکیبی نسبتاً ضعیفی برای پیش‌گویی زود هنگام دیابت بارداری محسوب می‌شود (شکل ۲). ذکر این نکته ضروری به نظر می‌رسد که با اضافه نمودن اثر پیشگوه‌های شاخص توده بدنی و سن نیز تغییر قابل ملاحظه‌ای در مقادیر معیارهای تشخیص حاصل نشد. از

جدول ۱. مقادیر معیارهای تشخیصی برای هر یک از نشانگرهای زیستی در اوایل سه ماهه دوم بارداری

مقدار p-	AUC	منفی اخباری ارزش (درصد)	مثبت اخباری ارزش (درصد)	ویژگی (درصد)	حساسیت (درصد)	برش نقطه بهینه	متغیر
۰/۱۰۳	۰/۵۳۱	۸۹/۹۰	۱۶/۸۱	۶۰	۵۳/۸۲	۰/۱۲	AFP (میلی لیتر بر نانوگرم)
۰/۰۰۱	۰/۵۹۷	۹۲/۳۰	۱۷/۸۰	۵۶/۲۳	۶۶/۶۶	۰/۱۱	uE3 (میلی لیتر بر نانوگرم)
۰/۰۸۱	۰/۵۶۷	۹۱/۱۴	۴۶/۰۲	۴۹/۲۹	۷۳/۰۱	۰/۱۱	β -HCG (mIU/ml)

جدول ۲. مقادیر معیارهای تشخیصی برای ترکیب‌های دوتایی نشانگرهای زیستی در اوایل سه ماهه دوم بارداری

مقدار p-	AUC	α	منفی اخباری ارزش (درصد)	مثبت اخباری ارزش (درصد)	ویژگی (درصد)	حساسیت (درصد)	برش نقطه بهینه	متغیر
۰/۰۰۷	۰/۵۹۳	۰/۳۴	۹۱/۵۱	۱۶/۲۵	۵۸/۳۰	۶۱/۹۰	۰/۱۱	uE3 + AFP
۰/۰۰۷	۰/۵۷۵	-۰/۱۰	۹۰/۱۵	۱۸/۶۶	۶۵/۱۵	۴۲/۶۸	۰/۱۲	AFP + β -HCG
۰/۰۲۱	۰/۵۸۶	۰/۰۴	۹۰/۹۷	۱۵/۶۴	۵۶/۱۴	۶۰/۳۱	۰/۱۲	β -HCG + uE3

جدول ۳. مقادیر معیارهای تشخیصی برای ترکیب‌های سه تایی نشانگرهای زیستی در اوایل سه ماهه دوم بارداری

مقدار p-	AUC	α	اخباری ارزش منفی (درصد)	اخباری ارزش مثبت (درصد)	ویژگی (درصد)	حساسیت (درصد)	برش نقطه بهینه	متغیر
۰/۰۰۴	۰/۵۹۳	۰/۱۴	۹۱/۶۴	۱۶/۷۴	۵۷/۵۵	۶۱/۹۰	۰/۱۲	AFP + uE3 + β -HCG
< ۰/۰۰۱	۰/۶۲۸	۰/۱۶	۹۱/۶۹	۲۱/۰۸	۶۵/۶۲	۵۳/۲۰	۰/۱۴	+ AFP + uE3 + β -HCG + سن
< ۰/۰۰۱	۰/۶۶۲	۰/۲۷	۹۳/۶۳	۲۰/۸۷	۶۴/۳۳	۶۸/۲۵	۰/۱۳	AFP + uE3 + β -HCG + بدنی توده شاخص
< ۰/۰۰۱	۰/۷۵۱	۰/۳۴	۹۵/۴۱	۲۸/۷۷	۷۴/۶۲	۸۲/۹۵	۰/۱۰	AFP + uE3 + β -HCG + بدنی توده شاخص + سن

بحث و نتیجه‌گیری

در پژوهش حاضر، برای بالا بردن دقت رده‌بندی در پیش‌بینی زودهنگام بیماری از منحنی راک ترکیبی با تعدیل اثر پیشگوهای مؤثر استفاده شد. این منحنی در زمانی کاربرد دارد که ارزش تشخیصی هر نشانگر زیستی به تنهایی و به اندازه کافی بالا نباشد. در ترکیب نشانگرها، مسئله مهم شناسایی روش‌های آماری مؤثری است که بتواند ترکیبی را با عملکرد رده‌بندی مناسب شناسایی کند. در این راستا محققان از رویکردهای مختلفی چون درخت تصمیم، شبکه عصبی و تصمیم بیزی استفاده کرده‌اند [۳۳، ۳۲]. به هر حال بسته به این‌که هدف، ماکزیم‌سازی AUC و یا بیشینه کردن حساسیت در دامنه‌ای از ویژگی‌ها باشد، روش‌ها با هم متفاوت خواهد بود. در حقیقت ممکن است روش‌هایی که برای یک هدف به‌کار می‌رود، برای هدف دیگر کارایی لازم را نداشته باشد. رگرسیون لجستیک یک روش معمول و البته شایع برای ترکیب نشانگرها است. این روش تکیه بر مشخصه‌سازی فرم پارامتریک مدل دارد. اما با این حال در زمان نقض فرض مدل، رگرسیون لجستیک از عملکرد طبقه‌بندی خیلی خوبی برخوردار نیست [۳۴]. پیه در سال ۲۰۰۶ استفاده

از AUC را به عنوان یک تابع هدف برای به حداکثر رساندن دقت تشخیص در هنگام ترکیب خطی مارکرها پیشنهاد داد. در واقع او با انجام شبیه‌سازی دریافت که روش‌های رگرسیونی مبتنی بر درست‌نمایی برای ترکیب مارکرها زمانی مناسب و از نظر آماری کارا هستند که مدل رگرسیونی صحیح و فرض‌ها برقرار باشد؛ در غیر این صورت ماکزیم‌سازی AUC به‌جای درست‌نمایی ممکن است ترجیح داده شود [۳۵].

در این مقاله، از روشی استفاده شد که یک فرم ترکیبی را که بر مبنای یک قانون تصمیم‌گیری خاص نیست، استخراج می‌کند. اما با این حال بهتر است چنین قانونی از منابع مختلف مانند اطلاعات درباره هزینه خطا در اندازه‌گیری و نیز شیوع بیماری استخراج شود. با داشتن چنین اطلاعاتی، می‌توان اندازه‌ای از فاصله بین توزیع‌های غرات خطی را برای دو رده سالم و بیمار ماکزیم کرد. به علاوه از آن‌جا که منحنی راک به عنوان معیاری قابل قبول در پزشکی تشخیصی به‌کار می‌رود، بنابراین طبیعی است که استفاده از AUC به عنوان تابع هدف برای بهینه‌سازی مناسب باشد. روش به‌کار رفته در این پژوهش تا حدودی شبیه به تحلیل جداسازی است. در تحلیل جداسازی نسبت واریانس

انتخاب شد. در واقع این ترکیب می‌تواند به‌طور معناداری وضعیت ۸۲ درصد مادرانی که به دیابت بارداری مبتلا می‌شوند را به درستی پیش‌بینی کند. سانکن و بارتلس در سال ۲۰۰۱ در تعیین رابطه این سه نشانگر با دیابت بارداری به نتیجه معناداری دست نیافتند [۴۰]. در مقابل راتی و همکاران رابطه معناداری را بین دیابت بارداری و AFP و β -HCG گزارش کردند [۲۶]. بر اساس نتایج تورنبرگ و همکاران AFP در پیش‌بینی دیابت بارداری به‌طور معناداری مؤثر بود [۴۱]. هور و همکاران نیز در سال ۲۰۱۷ نشان دادند uE3 و β -HCG پیشگوه‌های مناسبی برای دیابت بارداری در اوایل سه ماهه دوم بارداری هستند [۲۸]. تنها در یک مطالعه توسط ساین و همکارانش، معیارهای دقت این سه نشانگر (به صورت جداگانه) در تشخیص زودهنگام دیابت بارداری در سه ماهه دوم بارداری مورد بررسی قرار گرفتند. مقدار حساسیت، ویژگی و AUC برای AFP به ترتیب برابر با ۳۲/۳ درصد، ۷۸ درصد و ۰/۵۱۲ به‌دست آمد. برای HCG مقادیر شاخص‌ها به ترتیب برابر با ۶۹/۶ درصد، ۴۷/۹ درصد و ۰/۵۶۵ و برای uE3 برابر با ۳۶/۲ درصد، ۷۸/۵ درصد و ۰/۵۷۱ بود. از بین سه نشانگر فوق تنها HCG و uE3 با دیابت بارداری ارتباط معنادار داشتند [۴۲]. به‌طور کلی با مقایسه نتایج این مطالعه با یافته‌های پژوهش حاضر در می‌یابیم که ترکیب سه‌تایی قدرت پیش‌تری در تشخیص بیماری دارد. بنابراین استفاده از این ترکیب با توجه به استفاده از اطلاعات پیش‌تر و دقت بالاتر، در اوایل سه ماهه دوم بارداری (بین هفته ۱۴ تا ۱۷) برای پیش‌بینی زودهنگام دیابت پیشنهاد می‌شود.

دیابت بارداری یک نگرانی مهم اجتماعی و اقتصادی بوده و با توجه به افزایش کم‌تحرکی زنان باردار و بالا رفتن سن ازدواج و بارداری در بسیاری از کشورهای در حال توسعه و توسعه‌یافته یکی از مشکلات رو به رشد مامایی است. لذا در صورت بالا بودن قدرت تشخیصی uE3، AFP و β -HCG استفاده از آن‌ها برای پیش‌بینی زودرس دیابت مؤثر خواهد بود. با بررسی‌های انجام گرفته، به جرأت می‌توان گفت این پژوهش اولین مطالعه‌ای است که دقت هر سه تست غربالگری یاد شده را به صورت هم‌زمان و با تعدیل کردن اثر کووریت‌های مهم مورد ارزیابی قرار داده است. با توجه به یافته‌های به‌دست آمده به دست‌اندرکاران مراقبت‌های بهداشتی پیشنهاد می‌شود که بر اساس ترکیب نهایی انتخاب شده، غربالگری مناسب و اقدام‌های پیشگیرانه را جهت کاستن عوارض نامطلوب دیابت انجام دهند. البته این نکته را هم باید در نظر گرفت که ممکن است استفاده از ترکیب نشانگرهای زیستی بار اقتصادی را بر خانواده‌ها و مراکز درمانی تحمیل کند. در نتیجه به نظر می‌رسد اتکا به روش ترکیبی کافی نبوده و شاید بتوان از روش‌های دیگر جهت تشخیص زودهنگام دیابت

بین گروهی به واریانس درون گروهی ماکزیم می‌شود. سو و لیو به بحث در مورد این موضوع پرداختند و روش پیشنهادی خود را روی داده‌های نرمال چندمتغیره بسط دادند [۱۶]. در این مطالعه، از سه نشانگر زیستی استفاده شد. در حالت خاص و با داشتن تنها دو نشانگر، روش‌های محاسباتی به‌کار رفته برای ترکیب کردن نسبتاً ساده‌تر است. اما با افزایش تعداد نشانگرها ($K > 2$)، پیچیدگی‌های محاسباتی در یافتن ضریب $\alpha = \{\alpha_2, \dots, \alpha_K\}$ برای بیشینه کردن AUC بیش‌تر می‌شود. لیو و همکارانش روش توزیع آزاد min-max را پیشنهاد دادند که به‌طور خطی فقط مقادیر می‌نیم و ماکزیم مقادیر k نشانگر زیستی را برای ماکزیم کردن AUC با یک‌دیگر ترکیب می‌کند. اما ضعف این روش آن است که چون تنها از مقادیر ماکزیم و می‌نیم استفاده می‌کند، لذا ممکن است بهترین روش برای دستیابی به ترکیب خطی بهینه نباشد [۱۷]. کانگ و همکاران روش ترکیبی گام به گام را با استفاده از حجم زیر رویه راک به عنوان تابع هدف برای نشانگرهایی با مقیاس ترتیبی با روش ناپارامتری پیشنهاد دادند که همه مارکرها را به شیوه گام به گام با یک‌دیگر ترکیب می‌کند [۱۸]. در مقاله حاضر نیز از این رویکرد برای یافتن ترکیب بهینه استفاده شد. مزیت استفاده از این روش آن است که در هر گام از انجام محاسبات می‌توان از دو نشانگر زیستی در یک زمان استفاده کرد. اما با این وجود ممکن است $\{\alpha_2, \dots, \alpha_K\}$ یافت شده از این روش در فضای $K=2$ منجر به ایجاد یک ترکیب بهینه نشود. این موضوع مهمی است که در مطالعات آینده به لحاظ کاربردی بهتر است مورد بررسی قرار بگیرد.

در مطالعه حاضر، با توجه به نرمال بودن توزیع متغیرها از روش پارامتری برای ترکیب استفاده شد. افرادی چون رایدویک و همکاران [۳۶] و وو و همکاران [۳۷] نیز روش پارامتری را در برآورد معیارهای تشخیصی ترکیب تست‌ها به‌کار بردند. در مقابل محمد و همکارانش [۳۸] و یان و لیو [۳۹] تحت فرض غیر نرمال، دقت تشخیصی تست‌ها را در حالت ترکیبی مورد ارزیابی قرار دادند. در مطالعات یاد شده بالا رفتن ارزش تشخیصی تست‌ها به هنگام ترکیب گزارش شد که با یافته‌های این مقاله هم‌خوانی دارد. در این تحقیق ابتدا ترکیب‌های دوتایی نشانگرهای زیستی مد نظر قرار گرفت. از میان این ترکیب‌ها، ترکیب uE3 و AFP دارای بزرگ‌ترین مقدار AUC بود. اما به دلیل پایین بودن نسبی مقادیر شاخص‌های ارزیابی دقت تشخیصی، این ترکیب قدرت کافی برای تشخیص زودرس دیابت بارداری را نداشت. در ترکیب‌های سه‌تایی، ترکیبی که به‌طور هم‌زمان اثر سن و شاخص توده بدنی را به حساب آورده بود، به علت بالا بودن مقادیر معیارهای تشخیصی (حساسیت برابر با ۸۲/۹۵ درصد، ویژگی برابر با ۷۴/۶۲ درصد و $AUC = 0.751$) به عنوان بهترین ترکیب بهینه

[12] Kwak JY, Hwang H, Kim S-K, Choi JY, Lee SM, Bang H, et al. Prediction of sarcopenia using a combination of multiple serum biomarkers. *Sci Rep* 2018; 8: 8574.

[13] Etzioni R, Kooperberg C, Pepe M, Smith R, Gann PH. Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics* 2003; 4: 523-538.

[14] Guo S, Huang CC, Zhao W, Yang AC, Lin CP, Nichols T, Tsai SJ. Combining multi-modality data for searching biomarkers in schizophrenia. *PLoS One* 2018; 13: e0191202.

[15] Huang X, Qin G, Fang Y. Optimal combinations of diagnostic tests based on AUC. *Biometrics* 2011; 67: 568-576.

[16] Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *J Am Stat Assoc* 1993; 88: 1350-1355.

[17] Liu C, Liu A, Halabi S. A min-max combination of biomarkers to improve diagnostic accuracy. *Stat Med* 2011; 30: 2005-2014.

[18] Kang L, Xiong C, Crane P, Tian L. Linear combinations of biomarkers to improve diagnostic accuracy with three ordinal diagnostic categories. *Stat Med* 2013; 32: 631-643.

[19] Unal I. Defining an optimal cut-point value in roc analysis: an alternative approach. *Comput Math Methods Med* 2017; 2017: 3762651.

[20] Li C, Tao H, Yang X, Zhang X, Liu Y, Tang Y, et al. Assessment of a combination of Serum Proteins as potential biomarkers to clinically predict Schizophrenia. *Int J Med Sci* 2018; 15: 900-906.

[21] Schisterman EF, Faraggi D, Reiser B. Adjusting the generalized ROC curve for covariates. *Stat Med* 2004; 23: 3319-3331.

[22] Janes H, Longton G, Pepe M. Accommodating covariates in ROC analysis. *Stata J* 2009; 9: 17-39.

[23] Liu D, Zhou XH. Covariate adjustment in estimating the area under ROC curve with partially missing gold standard. *Biometrics* 2013; 69: 91-100.

[24] Bagheri R, Hekmat K, Abedi P, Tabesh H, Omidifar Z. Relationship of gestational diabetes with serum levels of retinol. *Koomesh* 2014; 15: 551-556. (Persian).

[25] Özkaya E, Çakır E, Çınar M, Altay M, Gelişen O, Kara F. Second trimester serum alpha-fetoprotein level is a significant positive predictor for intrauterine growth restriction in pregnant women with hyperemesis gravidarum. *J Turk Ger Gynecol Assoc* 2011; 12: 220-224.

[26] Rätty R, Anttila L, Virtanen A, Koskinen P, Laitinen P, Mörsky P, et al. Maternal midtrimester free β -HCG and AFP serum levels in spontaneous singleton pregnancies complicated by gestational diabetes mellitus, pregnancy-induced hypertension or obstetric cholestasis. *Prenat Diagn* 2003; 23: 1045-1048.

[27] Spandana T, Chaudhuri J, Silambanan S. Assessing the need for adjustment of first trimester screening markers in Diabetic Women. *IJCRR* 2015; 2: 190-193.

[28] Hur J, Cho EH, Baek KH, Lee KJ. Prediction of gestational diabetes mellitus by unconjugated estriol levels in maternal serum. *Int J Med Sci* 2017; 14: 123-127.

[29] Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol* 2018; 63: 07TR01.

[30] Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 2013; 4: 627-635.

[31] Feng Z, Yasui Y. Statistical considerations in combining biomarkers for disease classification. *Dis Markers* 2004; 20: 45-51.

[32] Erdman LK, Dhabangi A, Musoke C, Conroy AL, Hawkes M, Higgins S, et al. Combinations of host biomarkers predict mortality among Ugandan children with severe malaria: a retrospective case-control study. *PLoS One* 2011; 6: e17440.

[33] Duan X, Yang Y, Tan S, Wang S, Feng X, Cui L, et al. Application of artificial neural network model combined with four biomarkers in auxiliary diagnosis of lung cancer. *Med Biol Eng Comput* 2017; 55: 1239-1248.

[34] Fong Y, Yin S, Huang Y. Combining biomarkers linearly and nonlinearly for classification using the area under the ROC curve. *Stat Med* 2016; 35: 3792-3809.

[35] Pepe MS, Cai T, Longton G. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* 2006; 62: 221-229.

[36] Rydevik G, Innocent GT, Marion G, Davidson RS, White PC, Billinis C, et al. Using combined diagnostic test results to hindcast trends of infection from cross-sectional data. *PLoS Comput Biol* 2016; 12: e1004901.

بارداری استفاده نمود که از نظر انجام راحت تر و بار مالی کمتری داشته باشد.

در انتها لازم است به چند محدودیت این مطالعه اشاره شود. اول این که چون این مطالعه مقطعی بود، علیت عوامل مختلف را نمی توان توضیح داد. دوم این که کنترل برخی متغیرهای مخدوش کننده در زمان اجرا ممکن نبود. سوم این که چون جامعه آماری مورد مطالعه تنها شامل مادران بارداری بود که در این مطالعه ثبت نام کرده بودند و این مادران نماینده کل مادران باردار از نظر وضعیت ابتلا به دیابت بارداری نیستند، بنابراین یافته های به دست آمده قابل تعمیم به همه مادران مراجعه کننده به همه مراکز درمانی واقع در شهر تهران نیست. با توجه به محدودیت های یاد شده پیشنهاد می شود مطالعه ای در سطح وسیع تر روی همه مادران باردار از نظر وضعیت ابتلا به دیابت بارداری و کنترل عوامل مستقیم و غیر مستقیم خطر ساز دیابت بارداری به صورت هم زمان برای بالا بردن هر چه بیشتر قدرت تشخیص انجام شود.

تشکر و قدردانی

نویسندگان این مقاله از حمایت دانشگاه تربیت مدرس و نیز دانشگاه علوم پزشکی شهید بهشتی صمیمانه تشکر و قدردانی می نمایند.

منابع

[1] Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004; 50: 473-476.

[2] Bossuyt PM, Irwig L, Craig J, Glasziou P. Diagnostic accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006; 332: 1089-1092.

[3] Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; 336: 1106-1110.

[4] Naaktgeboren CA, De Groot JA, van Smeden M, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med* 2013; 159: 195-202.

[5] Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *J Eval Clin Pract* 2006; 12: 132-139.

[6] Cheam AS, McNicholas PD. Modelling receiver operating characteristic curves using Gaussian mixtures. *Comput Stat Data Anal* 2016; 93: 192-208.

[7] Haker S, Wells WM, Warfield SK, Talos IF, Bhagwat JG, Goldberg-Zimring D, et al. Combining classifiers using their receiver operating characteristics and maximum likelihood estimation. *Med Image Comput Comput Assist Interv* 2005; 8: 506-514.

[8] Leung Sf, Tam JS, Chan AT, Zee B, Chan LY, Huang DP, et al. Improved accuracy of detection of nasopharyngeal carcinoma by combined application of circulating Epstein-Barr virus DNA and anti-Epstein-Barr viral capsid antigen IgA antibody. *Clin Chem* 2004; 50: 339-345.

[9] Yuan Z, Ghosh D. Combining multiple biomarker models in logistic regression. *Biometrics* 2008; 64: 431-439.

[10] Ma S, Huang J. Combining multiple markers for classification using ROC. *Biometrics* 2007; 63: 751-757.

[11] Kang L, Liu A, Tian L. Linear combination methods to improve diagnostic/prognostic accuracy on future observations. *Stat Methods Med Res* 2016; 25: 1359-1380.

[41] Thornburg LL, Knight KM, Peterson CJ, McCall KB, Mooney RA, Pressman EK. Maternal serum alpha-fetoprotein values in type 1 and type 2 diabetic patients. *Am J Obstet Gynecol* 2008; 199: 135.

[42] Sayin NC, Canda MT, Ahmet N, Arda S, Süt N, Varol FG. The association of triple-marker test results with adverse pregnancy outcomes in low-risk pregnancies with healthy newborns. *Arch Gynecol Obstet* 2008; 277: 47-53.

[37] Wu K, Cheng Y, Li T, Ma Z, Liu J, Zhang Q, et al. The utility of HbA1c combined with haematocrit for early screening of gestational diabetes mellitus. *Diabetol Metab Syndr* 2018; 10: 14.

[38] Muhammad N, Coolen F, Coolen-Maturi T. Nonparametric predictive inference with parametric copulas for combining diagnostic tests. *Stat Optim Inf Comput* 2017; 6: 398-408.

[39] Yan L, Tian L, Liu S. Combining large number of weak biomarkers based on AUC. *Stat Med* 2015; 34: 3811-3830.

[40] Sancken U, Bartels I. Biochemical screening for chromosomal disorders and neural tube defects (NTD): is adjustment of maternal alpha-fetoprotein (AFP) still appropriate in insulin-dependent diabetes mellitus (IDDM)? *Prenat Diagn* 2001; 21: 383-386.

Application of adjusted-receiver operating characteristic curve analysis in combination of biomarkers for early detection of gestational diabetes mellitus

Maedeh Amini (Ph.D Candidate)¹, Anoshirvan Kazemnejad (Ph.D)^{*1}, Farid Zayeri (Ph.D)², Azam Amirian (M.Sc)³, Nourossadat Kariman (Ph.D)⁴

1- Dept. of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

2-Proteomics Research Center and Department of Biostatistics, Faculty of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

3- Dept. of Midwifery, School of Nursing and Midwifery, Jiroft University of Medical Sciences, Jiroft, Iran

4- Dept. of Midwifery and Reproductive Health, School of Nursing and Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran

* Corresponding author. +98 21-82883875 kazem_an@modares.ac.ir

Received:27 Oct 2018 ; Accepted: 30 Apr 2019

Introduction: In medical diagnostic field, evaluation of diagnostic accuracy of biomarkers or tests has always been a matter of concern. In some situations, one biomarker alone may not be sufficiently sensitive and specific for prediction of a disease. However, combining multiple biomarkers may lead to better diagnostic. The aim of this study was to assess the performance of combination of biomarkers to early detection of gestational diabetes.

Materials and Methods: In the present study, the information of 523 pregnant women referring to Mahdiah and Taleghani hospitals located in Tehran city was used. Relatively, Unconjugated Estriol (uE3), Alfa-Fetoprotein (AFP), and Beta- Human Chorionic Gonadotropin (β -HCG) were measured in the early second trimester of pregnancy. The accuracy of these biomarkers and also finding optimal linear combination of them was evaluated by area under receiver operating characteristic (ROC) curve (AUC), sensitivity, and specificity. Choosing of the best cut-point was based on Youden index. Data were statistically analyzed applying R 3.5.1 software package.

Results: In the combination of two biomarkers, the largest AUC, sensitivity, and specificity values were 0.593, 61.90%, and 58.30%, respectively with optimal cut-point value 0.11. In the combination of the three biomarkers by adjusting for age and BMI effects simultaneously, the largest AUC, sensitivity, and specificity values were 0.751, 82.95%, and 74.62%, respectively with optimal cut-point value 0.10.

Conclusion: Based on the findings of this research, the linear combination of the three biomarkers by considering covariate effects improved the diagnostic performance.

Keywords: Biomarker, Receiver Operating Characteristic Curve, Gestational Diabetes Mellitus, Early Diagnosis.