

بر اساس تصویب اداره کل آموزش مداوم جامعه پزشکی وزارت بهداشت، درمان و آموزش پزشکی به پزشکان عمومی، متخصصین رشته‌های جراحی، کلیه و مجاری ادراری و تناسلی، کودکان، آسیب‌شناسی تشریحی و بالینی، زنان و زایمان، جراحی استخوان و مفاصل، بیماری‌های اعصاب، پزشکی اجتماعی، بیماری‌های داخلی، بیماری‌های عفونی و گرمسیری، بیماری‌های قلب و عروق، بیهوشی، بیماری‌های پوست، جراحی عمومی، جراحی مغز و اعصاب، چشم پزشکی، پرتودرمانی، رادیولوژی، روانپزشکی، گوش و حلق و بینی، جراحی سر و گردن، پزشکی هسته‌ای، طب کار، پزشکی قانونی، پزشکی فیزیکی و توانبخشی، طب اورژانس، پزشکی ورزشی، دندانپزشکان عمومی و متخصصین رشته‌های مختلف دندانپزشکی که به حداقل ۷۰٪ از پرسش‌های مطرح شده در این مقاله پاسخ درست دهند ۱/۵ امتیاز تعلق می‌گیرد.

## ارزیابی دقت آزمون‌های تشخیصی

دکتر علی اکبر حق دوست<sup>۱\*</sup>، دکتر الهام ایرانمنش<sup>۲</sup>

### اهداف آموزشی

انتظار می‌رود بعد از مطالعه این متن خوانندگان، شاخص‌های سنجش دقت آزمون‌های تشخیصی را بشناسند. نحوه ترکیب آزمون‌های تشخیصی و مزایا و معایب آنها را بدانند. بتوانند تفسیر دقیق و کاربردی نتایج آزمون‌های تشخیصی را ارائه دهند.

۱-دانشیار اپیدمیولوژی، دانشکده بهداشت و مرکز تحقیقات فیزیولوژی، دانشگاه علوم پزشکی و خدمات بهداشتی - درمانی کرمان، ۲- پزشک عمومی، دفتر آموزش مداوم، دانشگاه علوم

پزشکی و خدمات بهداشتی - درمانی کرمان

\* نویسنده مسؤول، آدرس: معاونت آموزشی، دانشگاه علوم پزشکی کرمان • آدرس پست الکترونیک ahaghdoost@gmail.com

## مقدمه

تصور نمایید که یکی از بستگان شما به دلیل ضعف و بی‌حالی نسبتاً شدید از شما راهنمایی می‌خواهد. با کمی دقت متوجه می‌شوید که رنگ پریده است و با یک آزمایش ساده متوجه می‌شوید که مبتلا به کم‌خونی است. نگران می‌شوید و فکر می‌کنید که شاید خون‌ریزی مزمن باعث کم‌خونی وی شده است ولی در عین حال تشخیص‌های افتراقی دیگر مانند کم‌خونی به دلیل فقر آهن و البته بدخیمی نیز مد نظرتان می‌باشد. خوب فکر می‌کنید قدم بعدی برای نزدیک شدن به تشخیص نهایی چیست؟

بله باید آزمایش‌های تکمیلی بدهید. در اولین قدم فکر می‌کنید که شایع‌ترین محل خون‌ریزی مزمن از دستگاه گوارش است و برای همین سعی می‌کنید با یک آزمایش ساده مشخص نمایید که آیا در مدفوع خون مخفی وجود دارد یا خیر؟

اگر جواب این آزمایش مثبت شود آیا شما با قاطعیت می‌توانید بگوئید که خون‌ریزی مزمن از دستگاه گوارش وجود دارد و دلیل کم‌خونی وی همین خون‌ریزی است. آن روی دیگر سکه را نیز در نظر بگیرید، اگر جواب آزمایش منفی شد حتماً می‌توانید خون‌ریزی مخفی از دستگاه گوارش را رد نمائید؟

مسلماً جواب سؤالات فوق ساده نیست و وابسته به این است که آزمون انجام شده چه میزان دقت دارد. در این جا سعی خواهد شد تا به زبان ساده میزان دقت یک آزمون تشخیصی شرح داده شده و تحلیل منطقی نتایج آزمون بیان گردد.

## تعاریف پایه

اولین تعریف در بیان دقت یک آزمون تشخیصی، روش تشخیصی قطعی (gold standard) است. منظور از روش تشخیصی قطعی، آزمونی است که بدون خطا بوده و حرف نهایی را در تشخیص یک بیماری می‌زند. یعنی اگر برای فردی مثبت شد بتوان با یقین کامل فرد را بیمار و اگر منفی شد با یقین کامل بیماری را رد نمود. اما سؤال این است که

در صورت وجود چنین روش تشخیصی به چه دلیل ما به سمت روش‌های دیگر تشخیصی با دقت کمتر روی می‌آوریم؟

پاسخ این سؤال چندان دشوار نیست. اولاً اگر چنین روش تشخیصی وجود داشته باشد معمولاً بسیار گران و احتمالاً دشوار است و در هر شرایطی نمی‌توان آن را انجام داد و ثانیاً واقعاً در بسیاری از موارد چنین روشی وجود ندارد و فقط به صورت فرضی چنین آزمونی را تصور می‌کنیم. به عنوان مثال آیا شما روش تشخیصی قطعی برای سل ریوی سراغ دارید؟ جواب منفی است. حتی اگر نمونه‌برداری از بافت ریه را به عنوان روش تشخیصی قطعی معرفی نماییم باید بدانیم که ممکن است فرد، مبتلا به سل ریوی باشد ولی محلی که نمونه‌برداری شده است به هر دلیل عاری از باسیل سل باشد. پس منفی بودن آن دلیل بر رد حتمی بیماری نیست.

خوب حالا فرض نمایید که روش تشخیص قطعی برای یک بیماری وجود دارد ولی پرهزینه بوده و به راحتی نمی‌توان آن را انجام داد. برای همین از یک روش ساده‌تر برای تشخیص استفاده می‌نماییم و می‌خواهیم ببینیم چگونه می‌توان دقت تشخیصی روش ساده‌تر را محاسبه و بیان نمود.

دقت یک آزمون را با استفاده از ۳ شاخص اصلی حساسیت (sensitivity)، اختصاصی بودن (specificity) و صحت (accuracy) بیان می‌کنند. در ادامه به زبان ساده تعریف هر یک از این سه شاخص داده می‌شود.

حساسیت یعنی قدرت بیماریابی. به عبارت بهتر یعنی در چند درصد بیماران روش تشخیصی مورد نظر مثبت می‌شود. در مثال اول حساسیت آزمون نشان می‌دهد که روش مورد استفاده برای تشخیص خون مخفی در مدفوع چند درصد بیماران را بیمار تشخیص می‌دهد، یعنی اگر ۱۰۰ نفر که واقعاً خون‌ریزی گوارشی دارند را آزمایش نماییم در چند نفر از این ۱۰۰ نفر آزمایش مثبت می‌شود.

در مقابل اختصاصی بودن یعنی قدرت سالم‌یابی و بیان می‌کند که چند درصد افراد سالم نتیجه منفی خواهند داشت.

بیماران واقعی هستند که آزمایش آنها نیز مثبت می‌شود. به همین دلیل این افراد را مثبت واقعی (true positive) می‌نامیم. با توجه به این توضیحات فرمول حساسیت عبارت خواهد بود

$$\text{Sensitivity} = \frac{a}{a+c}$$

مشابه توجهات فوق را می‌توان برای نوشتن فرمول اختصاصی بودن بکار برد. قدرت سالم‌یابی یک آزمون نشان می‌دهد که چند درصد افراد واقعاً سالم (b+d) جواب آزمون منفی دارند (d). لذا فرمول آن به شرح زیر خواهد بود:

$$\text{Specificity} = \frac{d}{b+d}$$

بر این اساس افرادی که در خانه b قرار دارند کسانی هستند که واقعاً سالمند ولی آزمون در مورد آنها به صورت کاذب مثبت شده است لذا آنها را مثبت کاذب (false positive) نام‌گذاری می‌نماییم. در مقابل افراد بیماری که به اشتباه پاسخ آزمونشان منفی می‌شود در خانه c قرار می‌گیرند و لذا آنها را منفی کاذب (false negative) می‌نامیم.

فرمول صحت پاسخها نیز ساده خواهد بود. یعنی چند درصد پاسخها درست است. جوابهای درست بر اساس جدول فوق در خانههای a و d قرار می‌گیرند و فرمول آن عبارت خواهد بود:

$$\text{accuracy} = \frac{a+d}{a+b+c+d}$$

آیا به نظر شما برای درک این که دقت یک آزمون چه میزان است باید حتماً مقدار حساسیت و اختصاصی بودن را بدانیم یا دانستن میزان صحت به تنهایی کافی است؟ آیا اگر بگویم که صحت پاسخهای یک آزمون تشخیصی ۹۲٪ است شما قانع می‌شوید که حتماً حساسیت و اختصاصی بودن آن نیز بالا است؟

شاید پاسخ به این سؤال برایتان در این مرحله سخت باشد. بگذارید یک مثال عددی را با هم حل نماییم.

در مثال بالا یعنی اگر ۱۰۰ نفر که خونریزی گوارشی ندارند را مورد بررسی قرار دهیم، در چند درصد ایشان آزمون مورد نظر منفی خواهد شد.

در نهایت صحت نشان می‌دهد که چند درصد پاسخهای یک آزمون درست است. یعنی به نوعی چند درصد پاسخهای مثبت و منفی داده شده مورد اعتماد بوده و درست می‌باشند.

برای درک ساده‌تر موضوع بهتر است از یک جدول کوچک ولی بسیار کاربردی کمک بگیریم.

تبیجه آزمون	تشخیص نهایی یعنی پاسخ		تشخیص مورد نظر
	مثبت	منفی	
مثبت	a	b	a+b
	True positive	False positive	
منفی	c	d	c+d
	False negative	True negative	
کل	a+c	b+d	

لطفاً جدول فوق را به دقت مطالعه کرده و سعی نمایید معنی تک تک خانه‌های آن را درک کنید. در صورتی که درک خوبی از این جدول به دست آوردید باید به سرعت بگویید که حساسیت چگونه حساب می‌شود؟ به یاد داشته باشید که حساسیت یعنی قدرت بیماریابی، یعنی در چند درصد بیماران، آزمون مورد نظر مثبت می‌شود. بر این اساس بگویید بیماران در جدول فوق چگونه محاسبه می‌شوند؟ بیماران یعنی گروهی که تشخیص قطعی آنها مثبت است. پس جمع خانه‌های a و c بیماران واقعی را نشان می‌دهند. خوب حساسیت یعنی این که چند درصد این افراد که بیمار واقعی هستند در آزمون مورد نظر جواب مثبت خواهند داشت. درست است، افراد وارد شده در خانه a

بگوییم که در صورت منفی شدن آزمایش احتمال خونریزی گوارشی کم می‌شود ولی صفر نخواهد شد. برای محاسبه این که پاسخ مثبت و یا منفی به چه میزان می‌تواند دقت تشخیصی را تغییر دهد، بایست با دو اصطلاح ساده آشنا شویم. اصطلاح اول احتمال پیشین یا pretest است. قبل از انجام آزمایش بر اساس شرح حال و تجربه قبلی احتمال داشتن خونریزی گوارشی در فرد مورد نظر چقدر است؟ مثلاً می‌گویید من احتمال می‌دهم که فرد ۲۰٪ به این مشکل مبتلا باشد (و به عبارتی ۸۰٪ به این بیماری مبتلا نباشد). با این احتمال فرد آزمایش می‌دهد و جواب آزمایش وی مثبت می‌شود طبیعی است که فکر کنیم که احتمال بیمار بودن وی افزایش می‌یابد و مثلاً ۵۰٪ می‌شود. در مقابل اگر منفی شود احتمال سالم بودن فرد از ۸۰٪ به مثلاً ۹۰٪ می‌رسد. احتمالاتی که بعد از آزمایش به دست می‌آید را احتمال پسین یا posttest می‌نامیم که خود دو جزء دارد، ارزش اخباری مثبت (positive predictive value) یا همان ۵۰٪ و بیان می‌کند که اگر نتیجه آزمون مثبت شود احتمال بیمار بودن چقدر خواهد شد؛ با این تعریف ارزش اخباری منفی (negative predictive value) بیان می‌کند که اگر جواب آزمون منفی شد احتمال سالم بودن فرد چقدر است که در این مثال ۹۰٪ به دست آمد. در ادامه سعی می‌شود به زبان ساده روش محاسبه این اعداد بیان گردد. بدون شک ارزش‌های اخباری یک آزمون به میزان حساسیت و اختصاصی بودن یک آزمون مربوط است. اگر حساسیت آزمون مورد نظر در تشخیص خون منفی ۹۰٪ و اختصاصی بودن آن ۸۰٪ باشد و احتمال داشتن خونریزی گوارشی قبل از آزمایش در فرد ۲۰٪ باشد، می‌توان جدول قبلی را این چنین پر نمود.

نتیجه آزمون	تشخیص نهایی یعنی پاسخ		تشخیص مورد نظر
	gold standard		
	مثبت	منفی	جمع
مثبت	۵۰	۴۰	۹۰
منفی	۵۰	۸۶۰	۹۱۰
کل	۱۰۰	۹۰۰	۱۰۰۰

در جدول فوق مقدار حساسیت برابر ۵۰٪، مقدار اختصاصی بودن ۹۶٪ و مقدار صحت ۹۱٪ می‌باشد. بر این اساس نتیجه‌گیری شما چیست؟ می‌بینید در حالی که میزان صحت و میزان اختصاصی بودن بسیار بالا است ولی میزان حساسیت بسیار پایین است. پس باید نتیجه بگیریم که مقدار صحت بالا حتماً به معنای بالا بودن حساسیت و اختصاصی بودن نیست.

اما سؤال اول، آیا با دانستن میزان حساسیت و اختصاصی بودن آزمایش تشخیص خون مخفی می‌توانیم این عامل را قطعاً دلیل کم‌خونی بنامیم و یا آن را رد نماییم. مسلماً به این سادگی نمی‌توان قضاوت نمود. اگر در بیمار مورد نظر این آزمایش مثبت شد، می‌توان گفت که شاید واقعاً خونریزی مخفی وجود دارد و حال جواب آزمون نیز مثبت شده‌است، اما آیا احتمال دارد که این جواب به صورت کاذب مثبت شده‌باشد یعنی فرد مثلاً بعضی مواد غذایی را مصرف کرده باشد که در نتیجه، آزمایش باقیمانده این مواد غذایی در مدفوع را اشتباهاً خون تشخیص داده‌باشد؟ بله چنین احتمالی وجود دارد، پس به صرف مثبت شدن آزمون نمی‌توانیم مطمئن باشیم که فرد خونریزی مزمن دارد، ولی حداقل می‌توانیم بگوییم که احتمالاً خونریزی داشتن فرد با مثبت شدن جواب آزمایش بالا می‌رود، اما چقدر؟ این موضوعی است که در ادامه با هم بحث خواهیم نمود. در مقابل اگر جواب آزمایش منفی شده، باز نمی‌توانیم با قطع و یقین بگوییم که فرد خونریزی ندارد، بلکه احتمال منفی کاذب وجود دارد، یعنی ممکن است واقعاً خونریزی گوارشی باشد ولی آزمایش ما اشتباهاً منفی شود. اما باز می‌توانیم

همین توجیه چون اختصاصی بودن آزمون ۸۰٪ است لذا ۸۰٪ افراد سالم واقعاً سالم تشخیص داده خواهند شد که از ضرب کل تعداد ستون دوم (۸۰ نفر سالم) در اختصاصی بودن آزمون (یعنی ۸۰٪) عدد ۶۴ نفر منفی واقعی به دست خواهد آمد.

خوب برگردیم به تعریف ارزش اخباری مثبت. ارزش اخباری مثبت یعنی اگر پاسخ آزمون مثبت شد احتمال بیماری چقدر است؟ بر اساس جداول فوق، آزمون در چه کسانی مثبت می‌شود؟ بله درست گفتید در سطر اول جدول یعنی در ۳۴ نفر. از این تعداد چند نفر واقعاً بیمارند؟ ۱۸ نفر پس می‌توانیم بگوییم که از ۳۴ نفر مثبت ۱۸ نفر واقعاً بیمار و ۱۶ نفر مثبت کاذب هستند. لذا به راحتی می‌توانیم بگوییم که ارزش اخباری مثبت برابر است با حاصل تقسیم ۱۸ بر ۳۴

$$PPV = \frac{TP}{TP+FP} = \frac{18}{18+16} = 0.53 = 53\%$$

معنی این عدد این است که احتمال داشتن خونریزی گوارشی در فرد در صورت مثبت شدن جواب آزمایش از ۲۰٪ به ۵۳٪ افزایش می‌یابد.

آیا با این توضیحات شما می‌توانید ارزش اخباری منفی را محاسبه نمایید. یعنی باید ببینید در بین افرادی که جواب آزمایش منفی دارند چند درصد واقعاً سالم هستند

$$NPV = \frac{TN}{TN+FN} = \frac{64}{2+64} = 0.97 = 97\%$$

بر اساس نتایج فوق می‌توان نتیجه‌گیری کرد که احتمال داشتن خونریزی گوارشی در صورت مثبت بودن آزمون ۳۳٪ (از ۲۰٪ به ۵۳٪) و در صورت منفی بودن احتمال سالم بودن ۱۷٪ (از ۸۰٪ به ۹۷٪) افزایش می‌یابد. حال سؤال این است که عوامل موثر در افزایش دقت تشخیصی چه می‌باشند؟ مسلماً میزان حساسیت و اختصاصی بودن نتایج یک آزمون بسیار مهم است. هرچه آزمون مورد استفاده دقیق‌تر باشد بیشتر می‌توان به نتایج آن اعتماد نمود و لذا فاصله بین احتمال پیشین و پسین بیشتر خواهد شد. به نظر شما رابطه بین ارزش اخباری مثبت با حساسیت یک آزمون

نتیجه آزمون تشخیص مورد نظر	تشخیص نهایی داشتن خونریزی گوارشی بر اساس پاسخ gold standard		جمع
	مثبت	منفی	
مثبت	۱۸	۱۶	۳۴
منفی	۲	۶۴	۶۶
کل	۲۰	۸۰	۱۰۰

آیا می‌توانید بگویید که چرا در هر خانه این اعداد نوشته شده‌است و محاسبات مربوطه چگونه بوده‌است؟ برای راهنمایی جدول زیر را با دقت مطالعه نمایید.

نتیجه آزمون تشخیص مورد نظر	تشخیص نهایی داشتن خونریزی گوارشی بر اساس پاسخ gold standard		جمع
	مثبت	منفی	
مثبت	Sensitivity * (pretest)	۱۶	۳۴
منفی	Specificity * (1-pretest)	۲	۶۶
کل	Pretest	1-pretest	۱۰۰

شاید هنوز اعداد این دو جدول برایتان مفهوم نباشد. فکر کنید ۱۰۰ نفر مانند فرد مورد نظر با خصوصیات کاملاً مشابه دارید. خوب چون احتمال پیشین ۲۰٪ بود می‌توان تصور نمود که از این صد نفر بر اساس روش تشخیص قطعی باید ۲۰ نفر بیمار باشند یعنی باید جمع ستون اول ۲۰ باشد. از آن جایی که حساسیت آزمون مورد نظر ۹۰٪ بیان شده‌بود یعنی ۹۰٪ بیماران را بیمار تشخیص می‌دهد. به عبارتی از این ۲۰ نفر، ۱۸ نفر مثبت واقعی و در خانه بالا و سمت راست قرار می‌گیرند و ۲ نفر منفی کاذب می‌شوند. با

اشتباه کمتری افراد سالم را بیمار معرفی می‌کند. لذا اگر چنین آزمونی برای فردی مثبت شود احتمالاً به دلیل بیمار بودن واقعی فرد است نه به دلیل مثبت کاذب بودن جواب آن روش تشخیصی و بر این اساس ارزش اخباری مثبت آن بالا خواهد بود.

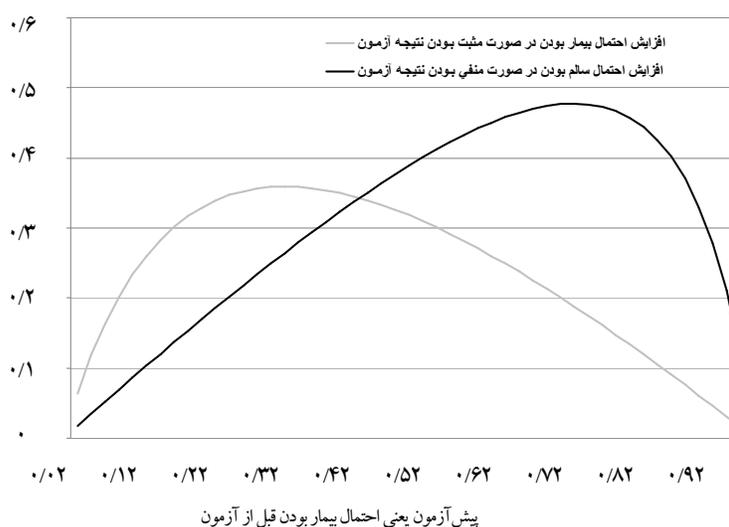
اگر چند پاراگراف بالا را خوب درک نموده باشید به راحتی خواهید توانست رابطه بین ارزش اخباری منفی و حساسیت را بیان نمایید. آزمون حساس منفی کاذب کمی دارد و در نتیجه اگر برای فردی منفی شود احتمالاً به دلیل سالم بودن فرد خواهد بود تا اشتباه تشخیصی به دلیلی منفی کاذب بودن.

اما عامل مهم دیگری که بر افزایش دقت تشخیصی تاثیرگذار است مقدار احتمال پیشین است. یعنی اگر این احتمال بالا رود میزان تاثیر پاسخ آزمون در تایید تشخیص و یا رد آن نیز افزایش می‌یابد ولی این رابطه خطی نیست. به نمودار زیر توجه نموده و سعی کنید معنی آن را به دقت درک کنید.

بیشتر است یا با اختصاصی بودن آن. به عبارتی اگر بخواهیم که ارزش اخباری مثبت بالا به دست بیاوریم باید از آزمونی استفاده کنیم که حساسیت بالا دارد یا آزمونی که اختصاصی‌تر است؟

برای پاسخ به فرمول ارزش اخباری مثبت نگاه کنید. مشاهده می‌کنید که در مخرج کسر هر چه قدر مقدار مثبت کاذب زیاد شود، مقدار ارزش اخباری مثبت کاهش می‌یابد. پس شاید مهم‌ترین عنصر در ارزش اخباری مثبت مقدار مثبت کاذب است و برای بالا بردن ارزش اخباری مثبت باید مقدار مثبت کاذب را کاهش داد. حالا فرمول‌های حساسیت و اختصاصی بودن را مرور نمایید. مشاهده خواهید کرد که در مخرج فرمول اختصاصی بودن مقدار مثبت کاذب وجود دارد و با افزایش اختصاصی بودن مقدار مثبت کاذب کاهش می‌یابد. در نتیجه می‌توان گفت که با بالا رفتن اختصاصی بودن یک آزمون مقدار مثبت کاذب کاهش و در نتیجه ارزش اخباری مثبت افزایش می‌یابد.

معنی این نتیجه‌گیری چیست؟ اگر آزمونی اختصاصی باشد یعنی اکثر افراد سالم را سالم تشخیص می‌دهد و لذا با



رابطه بین احتمال بیماری قبل از انجام آزمون و میزان تاثیر نتیجه آزمون در افزایش دقت تشخیصی، این نمودارها برای آزمون فرضی با حساسیت 90٪ و اختصاصی بودن 80٪ رسم شده

است.

چه باید کرد. ساده‌ترین کار آن است که ابتدا از آزمون‌های ساده و کم‌خرج استفاده نمود و سپس در زمانی که احتمال بیماری به حد قابل قبولی رسید از روش‌های گران قیمت استفاده کرد.

البته آزمون‌های تشخیصی ساده ممکن فقط یک سؤال و یا یک معاینه فیزیکی باشد. تصور نمایید که بیماری از کم‌درد رنج می‌برد. مسلماً یکی از تشخیص‌های افتراقی فتق دیسک کمر است. اما به صورت کلی این احتمال در قدم اول شاید بیش از ۳٪ نباشد و لذا توصیه به گرفتن عکس MRI با وجود داشتن دقت بالا چندان نمی‌تواند شانس بیماری مذکور را بالا ببرد. اگر از بیمار سؤال نماییم که آیا این درد به ران و یا پای وی تیر می‌کشد و جواب مثبت باشد شانس بیماری مذکور افزایش می‌یابد. این سؤال مانند یک آزمون تشخیصی عمل می‌کند و دارای حساسیت و اختصاصی بودن است که ممکن است در مقایسه با MRI زیاد بالا نباشد ولی خوب انجام آن نیز بسیار ساده و بدون هزینه است. جواب مثبت این سؤال ارزش اخباری مثبتی دارد که می‌توان آن را محاسبه نمود. مثلاً فکر نمایید که ارزش اخباری مثبت آن ۱۰٪ می‌باشد. به عبارتی شانس فتق دیسک کمر داشتن فردی که درد کمرش به ران و یا پای وی تیر می‌کشد از ۳٪ به ۱۰٪ می‌رسد. حال اگر معاینه فیزیکی انجام و نتیجه آن مثبت شود، شانس بیماری مذکور از ۱۰٪ به مثلاً ۳۰٪ افزایش می‌یابد. در این شرایط اگر عکس برداری MRI صورت گیرد، دقت تشخیصی به شدت افزایش می‌یابد و جواب مثبت و منفی آن برای تصمیم‌گیری بالینی تأثیرگذار خواهد بود.

در مثال فوق از روش سریال برای رسیدن به تشخیص نهایی استفاده شد. ابتدا شرح حال، سپس معاینه فیزیکی و در انتها عکس‌برداری. حالاً این سؤال می‌تواند مطرح شود که

بر اساس نمودار فوق اگر میزان پیش‌آزمون افزایش یابد، یعنی احتمال بیمار بودن فرد قبل از انجام آزمون افزایش یابد، تأثیر نتیجه مثبت آزمون در افزایش دقت بیماری به صورت تصاعدی افزایش می‌یابد به شکلی که در صورتی که قبل از آزمون، احتمال بیمار بودن حدود ۳۰٪ باشد نتیجه مثبت می‌تواند احتمال بیمار بودن را حدود ۳۵٪ افزایش دهد و در مجموع ارزش اخباری مثبت آزمون حدود ۶۵٪ خواهد بود. در مقابل اگر احتمال بیمار بودن قبل از آزمون ۷۵٪ باشد و به عبارتی احتمال سالم بودن ۲۵٪ باشد، پاسخ منفی می‌تواند احتمال سالم بودن را ۴۸٪ افزایش داده و در مجموع ارزش اخباری منفی حدود ۷۵٪ بوجود آورد.

در نمودار فوق کاملاً روشن است که اولاً ارزش‌های اخباری برخلاف میزان حساسیت و اختصاصی بودن کاملاً وابسته به میزان احتمال بیماری قبل از آزمون هست و بهتر است برای بالا بردن بهره حاصل از نتیجه آزمون در افزایش دقت تشخیصی، زمانی که شانس بیماری بسیار کم و یا بسیار زیاد است کمتر استفاده شود. ثانیاً با توجه به این که میزان شیوع بیماری‌ها در جوامع مختلف فرق می‌کند نمی‌توان ارزش اخباری یک آزمون در یک جامعه را به جامعه دیگر تعمیم داد. لذا بدون ترس از دوباره کاری می‌توان ارزش تشخیصی یک آزمون را در چند جامعه بررسی نمود حتی اگر حساسیت و اختصاصی بودن آن آزمون در همه جوامع ثابت باشد.

#### ترکیب سریال آزمون‌های تشخیصی

بر اساس این تفسیر انجام آزمون‌های بسیار گران و دقیق در زمانی که شانس بیماری کم است به اندازه کافی نمی‌تواند دقت نهایی تشخیص را افزایش دهد. در این موارد

$$PPV = \frac{TP}{TP+FP} = \frac{21}{21+242} = 8\%$$

حال اگر بخواهیم برای همه آنهایی که در معاینه فیزیکی پاسخ مثبت بوده‌اند، عکس برداری نماییم، جدول به شکل زیر پر خواهد شد:

نتیجه عکس- برداری	تشخیص نهایی داشتن فتق دیسک کمر بر اساس نتیجه gold standard		جمع
	مثبت	منفی	
مثبت	۲۰	۲۴	۴۴
منفی	۱	۲۱۸	۲۱۹
کل	۲۱	۲۴۲	۲۶۳

$$PPV = \frac{TP}{TP+FP} = \frac{20}{20+24} = 54\%$$

آیا متوجه شدید که جدول آخر چرا به این شکل پر شده‌است؟

زیاد سخت نیست، جدول اول نتایج معاینه فیزیکی را نشان می‌دهد. سطر بالای این جدول مربوط به کسانی است که معاینه فیزیکی مثبت دارند که بعضی از آنها واقعاً بیمارند و تعدادی از آنها مثبت کاذب می‌باشند و در مجموع ۲۶۳ نفر از ۱۰۰۰ نفر احتمالاً در معاینه فیزیکی مثبت خواهند شد. اگر دقت نماید عیناً سطر بالای جدول اول به عنوان سطر کل در جدول دوم وارد شده‌است. چرا که می‌خواهیم همه افرادی که در معاینه فیزیکی مثبت هستند را مورد عکس برداری قرار دهیم و بر اساس اطلاعات جدول اول، از این تعداد ۲۶۳ نفر، ۲۱ نفر واقعاً مبتلا به فتق دیسک کمر هستند و ۲۴۲ نفر مبتلا نیستند. نکته غامض در پر

آیا اگر توالی این آزمون‌ها عوض شود و مثلاً اول عکس- برداری و سپس معاینه و در انتها شرح حال گرفته‌شود آیا دقت نهایی تغییر خواهد کرد. بگذارید پاسخ این سؤال را با حل یک تمرین بدهم.

در مثال فوق برای ساده شدن کار، به جای سه آزمون از دو آزمون معاینه فیزیکی و عکس برداری استفاده می‌کنیم. قبل از هر آزمون شانس فتق دیسک کمر ۳٪ است. حساسیت و اختصاصی بودن معاینه فیزیکی برای تشخیص فتق دیسک کمر به ترتیب ۷۰٪ و ۷۵٪ است و این اعداد برای عکس برداری به روش MRI به ترتیب ۹۵٪ و ۹۰٪ است. اگر قیمت معاینه فیزیکی ۵۰۰۰ تومان و قیمت عکس برداری ۵۰۰۰۰ تومان باشد و بخواهیم در یک گروه ۱۰۰۰ نفری مبتلا به کمردرد از این دو روش تشخیصی استفاده نمایم چه باید بکنیم. اگر اول معاینه و در افراد مشکوک عکس برداری کنیم دقت نهایی چقدر خواهد بود و چه میزان باید هزینه کنیم؟ اگر معکوس عمل نماییم و ابتدا عکس برداری و سپس افراد مثبت را معاینه کنیم چه میزان هزینه خواهیم نمود؟

در حالت اول یعنی ابتدا معاینه فیزیکی و سپس عکس برداری، جدول مربوط به معاینه فیزیکی و ارزش اخباری مثبت آن عبارت خواهد بود از:

نتیجه معاینه فیزیکی	تشخیص نهایی داشتن فتق دیسک کمر بر اساس نتیجه gold standard		جمع
	مثبت	منفی	
مثبت	۲۱	۲۴۲	۲۶۳
منفی	۹	۷۲۸	۷۳۷
کل	۳۰	۹۷۰	۱۰۰۰

$$PPV = \frac{TP}{TP+FP} = \frac{28/5}{28/5+98} = 23\%$$

تشخیص نهایی داشتن فتق			
تبیجه معاینه	دیسک کمر بر اساس نتیجه جمع		فیزیکی
	gold standard		
	مثبت	منفی	
مثبت	۲۰	۲۴	۴۴
منفی	۸/۵	۷۳	۸۱/۵
کل	۲۸/۵	۹۷	۱۲۵/۵

$$PPV = \frac{TP}{TP+FP} = \frac{20}{20+24} = 45\%$$

مشاهده می‌فرمایید که در این حالت نیز مجدد به ارزش اخباری ۴۵٪ می‌رسیم که مشابه قبل است. به این معنی که انجام معاینه فیزیکی و سپس عکس برداری و یا ابتدا عکس برداری و سپس معاینه فیزیکی ما را به یک ارزش اخباری مثبت مساوی می‌رساند. در این مرحله به شما توصیه جدی می‌کنم که برای عمیق شدن درک مطلب همین محاسبه را برای ارزش اخباری منفی انجام دهید و ببینید که اگر نتیجه هر دو آزمون تشخیصی برای افراد منفی گردد ارزش اخباری منفی به چه عددی خواهد رسید و آیا این ارزش اخباری منفی وابسته به توالی آزمون‌ها هست یا خیر؟ چون در مثال قبل احتمال بیمار بودن ۳٪ و احتمال سالم بودن ۹۷٪ درصد است لذا ارزش اخباری منفی چندان نمایان نخواهد شد و برای این تمرین پیشنهاد می‌کنم که درصد بیماری را در ابتدا به جای ۳٪، ۳۰٪ در نظر بگیرید. اما بگذارید محاسبه هزینه‌ها را برای توالی دوم نیز حساب کنیم.

$$1000 \times 5000 + 125/5 \times 5000 = 18150000 = 50627500$$

نمودن جدول دوم دقیقاً همین نکته است و بقیه موارد عیناً شبیه جداول قبل است و شما با ضرب نمودن حساسیت و اختصاصی بودن در اعداد سطح کل می‌توانید خانه‌های جدول را پر نمایید.

بعد از این محاسبات مشخص می‌شود که در کل اگر در فردی که معاینه فیزیکی مثبت داشته، آزمون دوم یعنی عکس برداری نیز مثبت شود، احتمال داشتن فتق دیسک کمر به ۴۵٪ می‌رسد. خوب برای رسیدن به این دقت تشخیصی در این جامعه چه میزان باید هزینه نماییم؟

$$1000 \times 5000 + 263 \times 50000 = 18150000$$

بله درست محاسبه نمودید باید از همه افراد معاینه فیزیکی انجام شود که هر معاینه ۵۰۰۰ تومان هزینه در برخواهد داشت. سپس از همه افرادی که در معاینه مثبت تشخیص داده شده بودند (یعنی ۲۶۳ نفر) عکس برداری صورت گیرد که هزینه هر یک عکس ۵۰۰۰۰ تومان تعیین شده بوده است. به این شکل در کل باید بیش از ۱۸ میلیون تومان پول تشخیص پرداخت گردد تا به ارزش اخباری ۴۵٪ برسیم.

حالا محاسبه را معکوس نمایید و اول عکس برداری و سپس معاینه فیزیکی را مد نظر قرار دهید. جداول مذکور به شرح زیر خواهند شد:

تشخیص نهایی داشتن فتق			
نتیجه عکس برداری	دیسک کمر بر اساس نتیجه جمع		تبیجه معاینه
	gold standard		
	مثبت	منفی	
مثبت	۲۸/۵	۹۷	۱۲۵/۵
منفی	۱/۵	۸۷۳	۸۷۴/۵
کل	۳۰	۹۷۰	۱۰۰۰

افراد دیابتی وجود دارند که قند خون ناشتای آنها بین ۱۲۰ تا ۱۲۶ باشد، در حالت دوم این افراد مثبت تلقی می‌شوند و لذا قدرت بیماریابی افزایش و خطای منفی کاذب کم می‌شود. در عوض افراد سالمی هم که قند خون بین ۱۲۰ تا ۱۲۶ دارند بیمار گزارش شوند و در نتیجه مثبت کاذب افزایش و اختصاصی بودن کاهش می‌یابد. پس با کاهش نقطه برش میزان حساسیت به تدریج زیاد و مقدار اختصاصی بودن کاهش می‌یابد. با همین توجیه می‌توان دریافت که با افزایش خط برش نتیجه معکوس به دست آمده و میزان حساسیت کاهش و اختصاصی بودن افزایش می‌یابد.

برای تعیین بهترین نقطه برش باید عددی را انتخاب نماییم که حداکثر دقت را ایجاد می‌کند و به عبارتی جمع حساسیت و اختصاصی بودن آن به حداکثر می‌رسد. در این صورت جمع خطاهای مثبت و منفی کاذب به حداقل خواهد رسید. معمولاً برای نشان دادن این موضوع از نموداری بنام ROC (Receiver Operating Characteristic) استفاده می‌شود. در این نمودار محورهای عمودی و افقی به ترتیب مقدار حساسیت و (اختصاصی بودن-۱) را نمایش می‌دهد. می‌توان برای هر نقطه برش مقدار حساسیت و اختصاصی بودن را محاسبه نمود و سپس اعداد مذکور را در نمودار مشخص و به هم وصل کرد. در ادامه شکل شماتیکی از نمودار ROC آورده شده است. بهترین خط برش به صورت کلی نقطه‌ای از نمودار است که کمترین فاصله را تا نقطه یک بر روی محور  $y$  داشته باشد. البته روش آسانتر آن است که در یک جدول حساسیت و اختصاصی بودن نقاط برش مختلف محاسبه گردد و نقطه‌ای که بیشترین مقدار جمع این دو شاخص را ارائه می‌دهد به عنوان بهترین نقطه برش معرفی شود.

بله در این حالت باید بیش از ۵۰ میلیون تومان هزینه نماییم. مسلم است که هزینه‌ها در توالی اول بسیار کمتر است چراکه در حالت اول برای هر هزار نفر یک آزمون ارزان قیمت انجام شده و سپس آزمایش گران قیمت فقط برای تعداد محدودی که در آزمون اول مثبت شده‌اند انجام شده است ولی در حالت دوم عکس برداری که هزینه‌بر است برای همه انجام شده و لذا هزینه‌های کل را به شدت افزایش داده است.

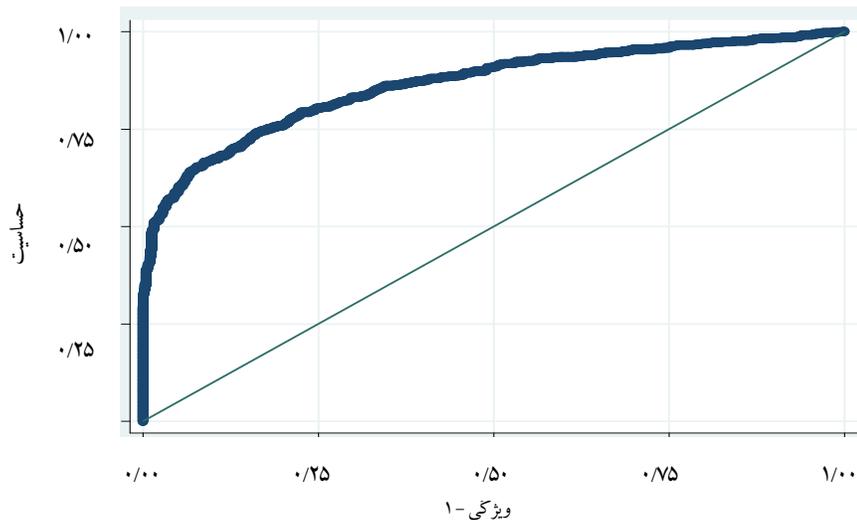
از مجموعه توضیحات فوق می‌توان نتیجه گرفت که برای رسیدن به یک قدرت تشخیصی بالا باید از ترکیب آزمون‌ها استفاده شود و دقت نهایی وابسته به توالی آزمون‌ها نیست ولی قیمت نهایی کار کاملاً وابسته به این است که ابتدا آزمون‌های ساده و ارزان و سپس آزمون‌های گران قیمت انجام پذیرد.

#### تعیین نقطه برش با حداکثر دقت تشخیصی

در بسیاری از آزمون‌های تشخیصی، نتیجه آزمایش به صورت یک عدد گزارش می‌شود که لازم است فرد تفسیر کننده بر اساس این که مقدار مذکور از یک حد نصاب بالاتر است یا خیر نتیجه را مثبت یا منفی اعلام نماید. آیا مثالی در این خصوص به ذهن تان می‌رسد؟

پاسخ آسان است، تعریف دیابت چیست؟ قند خون بالای ۱۲۶. تعریف فشارخون بالا چیست؟ فشارخون دیاستولیک بالای ۸۰ و بسیار از موارد مشابه. اما این نقاط برش از کجا به دست آمده‌اند؟ دلیل این که در تشخیص سل، مقدار PPD بالای ۱۰ میلی‌متر مثبت فرض می‌شود چه می‌باشد؟

فرض نمایید که به جای خط برش ۱۲۶ در تشخیص دیابت، مقدار را ۱۲۰ قرار دهیم، چه اتفاقی می‌افتد؟ احتمالاً



از تشخیص اشتباه بیماری (مثبت کاذب) با هزینه حاصل از عدم تشخیص بیماری به اشتباه (منفی کاذب) برابر است؟ مسلماً چنین نیست. مثلاً فکر کنید که اگر ما یک فرد مبتلا به بدخیمی را اشتبهاً تشخیص ندهیم مسلماً هزینه سنگینی به فرد و خانواده و جامعه تحمیل خواهد نمود و هزینه درمان وی در مراحل پیشرفته‌تر بیشتر و عمر مفید وی کاهش می‌یابد در مقابل هزینه مثبت کاذب اگرچه در کوتاه مدت به دلیل فشار روانی زیاد به فرد مبتلا و خانواده زیاد است ولی در مقایسه با هزینه منفی کاذب کمتر می‌باشد. پس بر این اساس باید جمع حساسیت و اختصاصی بودن تعدیل گردد. با این توضیحات می‌توان گفت که باید مقدار عبارت زیر به حداقل خود برسد.

$$(\text{Rate of FN}) \times (\text{FP cost}) + (\text{FN cost}) \times (\text{Rate of FP})$$

و یا کمی دستکاری می‌توان نوشت که باید مقدار عبارت زیر به حداکثر خود برسد.

$$(\text{Specificity}) \times (\text{FP cost}) + (\text{Sensitivity}) \times (\text{FN cost})$$

نمودار بالا یک نمونه خوب از منحنی ROC است. با استفاده از این نمودار همچنین می‌توان فهمید که آیا در کل این روش تشخیصی دقت کافی در تشخیص بیماری را دارد یا خیر؟ معمولاً در این نمودار سطح زیر منحنی را نیز حساب می‌کنند. هر چه سطح زیر منحنی از ۰/۵ بیشتر شود (خط رفرنس که در واقع قطر نمودار است) اعتبار تشخیصی روش مورد نظر بیشتر خواهد شد. در مثال فوق سطح زیر منحنی حدود ۰/۸۳ محاسبه شده است.

اگر دو آزمون تشخیصی برای یک موضوع وجود داشته‌باشد، آزمونی که منحنی ROC آن بالاتر و سمت چپ‌تر قرار گیرد سطح زیر منحنی بیشتری خواهد داشت و لذا در کل می‌توان گفت که ارزش تشخیصی بیشتری نیز به خود اختصاص خواهد داد.

اما نکته آخر این که همیشه نقطه برش، نقطه‌ای نیست که بیشترین حاصل جمع حساسیت و اختصاصی بودن را ایجاد می‌نماید. برای درک بهتر فکر کنید آیا هزینه حاصل

## سؤالات آموزش مداوم

۱. حساسیت یک آزمون ۹۰٪ است. این عبارت یعنی:
- الف - ۹۰٪ بیماران را تشخیص می‌دهد  
 ب - ۹۰٪ افراد سالم را تشخیص می‌دهد  
 ج - اگر مثبت شود احتمال بیمار بودن فرد ۹۰٪ است  
 د - اگر منفی شود احتمال سالم بودن فرد ۹۰٪ است
۲. برای بالا بردن دقت نهایی تشخیص کدام راهکار مناسب نیست؟
- الف - استفاده از آزمون‌های دقیق‌تر  
 ب - ترکیب موازی آزمون‌های تشخیصی  
 ج - ترکیب سریال آزمون‌های تشخیصی  
 د - استفاده از آزمون‌های گران‌تر و جدیدتر
۳. اگر اختصاصی بودن یک آزمون بالا باشد یعنی:
- الف - در رد نهایی بیماری باید بکار گرفته‌شود  
 ب - در تایید نهایی بیماری باید بکار گرفته‌شود  
 ج - دارای دقت بالا خواهد بود  
 د - بسته به شرایط گاه در رد و گاه در تایید بیماری می‌تواند مفید باشد
۴. بهترین شکل ترکیب آزمون‌های تشخیصی برای به دست آوردن بالاترین ارزش اخباری مثبت به کدام شکل طراحی باید طراحی گردد؟
- الف - ابتدا آزمون غیر دقیق و سپس آزمون دقیق  
 ب - ابتدا آزمون اختصاصی و سپس آزمون حساس  
 ج - ابتدا آزمون حساس و سپس آزمون اختصاصی  
 د - ترتیب آزمون‌ها در دقت نهایی تاثیری ندارد
۵. تاثیر نتیجه یک آزمون تشخیصی در افزایش ارزش اخباری به چه عواملی وابسته نیست؟
- الف - حساسیت آزمون  
 ب - اختصاصی بودن آزمون  
 ج - حجم نمونه جمعیت مورد مطالعه  
 د - احتمال بیماری قبل از انجام آزمون
۶. به چه دلیل بهتر است ابتدا آزمون‌های ساده و سپس آزمون‌های پیچیده‌تر را انجام داد؟
- الف - برای کاهش بار هزینه‌ها  
 ب - برای افزایش دقت تشخیصی  
 ج - برای کاهش هزینه‌ها و افزایش دقت تشخیصی  
 د - اهمیتی ندارد
۷. در صورتی که حد نصاب تشخیص دیابت با استفاده از قند خون ناشتا کاهش یابد، کدام حالت رخ نخواهد داد؟
- الف - حساسیت کم می‌شود  
 ب - اختصاصی بودن کم می‌شود  
 ج - ارزش اخباری مثبت کم می‌شود  
 د - ارزش اخباری منفی زیاد می‌شود
۸. برای به دست آوردن بهترین نقطه برش در تشخیص بیماری کدام عامل اهمیت ندارد؟
- الف - قیمت آزمون  
 ب - هزینه تحمیل شده در اثر تشخیص اشتباه بیماری  
 ج - هزینه تحمیل شده در اثر عدم تشخیص بیماری  
 د - رابطه بین حساسیت و اختصاصی بودن بیماری با تغییر نقطه برش

### قابل توجه شرکت کنندگان در برنامه خودآموزی:

شرکت کنندگان در برنامه خودآموزی لازم است فرم ثبت نام را به طور کامل تکمیل و به مهر نظام پزشکی ممهور نمایند و پس از مطالعه مقاله خودآموزی بعد از پاسخگویی به سؤالات پرسشنامه و اعلام نظر خود در خصوص مقاله مطالعه شده در فرم نظرخواهی نسبت به ارسال اصل هر سه فرم تکمیل شده حداکثر تا ۱۳۸۸/۶/۱۶ به آدرس کرمان، بلوار جمهوری اسلامی، ساختمان مرکزی دانشگاه علوم پزشکی، طبقه سوم، معاونت آموزشی، دفتر آموزش مداوم پزشکی، تلفن: ۰۳۴۱-۳۲۰۵۶۰۶ اقدام نمایند تا در صورت پاسخگویی صحیح به حداقل ۷۰٪ از سؤالات مقاله، گواهینامه شرکت در برنامه خودآموزی صادر و به آدرس مندرج در فرم ثبت نام ارسال گردد.

بسمه تعالی

جمهوری اسلامی ایران

وزارت بهداشت، درمان و آموزش پزشکی

معاونت آموزشی - اداره کل آموزش مداوم جامعه پزشکی

#### فرم ثبت نام در برنامه خودآموزی

عنوان مقاله: ارزیابی دقت آزمون‌های تشخیصی  
نام خانوادگی: نام پدر: نام: جنس: مرد  زن   
تاریخ تولد: محل فعالیت: استان: شهرستان: بخش: روستا:  
نوع فعالیت: هیأت علمی  آزاد  رسمی  پیمانی  قراردادی  طرح  سایر   
مقطع آخرین مدرک تحصیلی و سال اخذ مدرک:  
رشته تحصیلی مقاطع: لیسانس: فوق لیسانس: دکتر: تخصص: فوق تخصص:  
آدرس دقیق پستی: کد پستی: شماره تلفن: تاریخ تکمیل و ارسال فرم:  
امضاء، شماره نظام پزشکی و مهر متقاضی:

امضاء و مهر مسئول ثبت نام

سؤال	پاسخ	الف	ب	ج	د
۱					
۲					
۳					
۴					
۵					
۶					
۷					
۸					

نظری ندارم	کاملاً مخالقم	تأخیدی مخالقم	تأخیدی موافقم	کاملاً موافقم	خواهشمند است نظر خود را با گناردن علامت (X) در زیر گزینه مربوطه اعلام نمایید.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۱- محتوای مقاله بر اساس منابع جدید علمی ارائه شده است.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۲- محتوای مقاله با نیازهای حرفه‌ای من تناسب داشته است.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۳- محتوای مقاله در جهت تحقق اهداف آموزشی نوشته شده است.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴- در محتوای مقاله شیوایی و سهولت بیان در انتقال مفاهیم رعایت شده است.
سه عنوان پیشنهادی خود را برای ارائه مقالات خودآموزی ذکر نمایید.					
همکار گرامی لطفاً با ارائه نظرات و پیشنهادات خود در جهت توسعه کیفی مقالات خودآموزی، برنامه‌ریزان و مجریان برنامه‌های آموزش مداوم را یاری فرمایید.					