مقاله پژوهشی

فصلنامهٔ بیماری‌های پستان ایران، سال هشتم، شماره سوم، پاییز ۱۳۹۴

پیاده سازی سیستم تصمیم یار مبتنی بر شبکه عصبی احتمالی جهت تشخیص نوع سرطان پستان

آسیه خسروانیان*:گروه علمی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، ایران
سید سعید آیت: گروه علمی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، ایران

Archive of SID

چکیده

مقدمه: سرطان پستان رایج‌ترین شکل سرطان در زنان است. تشخیص به‌موقع سرطان شانس زنده ماندن بیمار را افزایش می‌دهد. شبکه‌های عصبی مصنوعی از روش‌های نوین مدل‌سازی و پیش‌بینی هستند. هدف از این مطالعه تشخیص خوش‌خیم یا بدخیم بودن توده‌های سرطان پستان است که برای این منظور سیستم تصمیم‌یار مبتنی بر شبکه عصبی احتمالی طراحی شد.

روش بررسی: در این مطالعه یک شبکه عصبی احتمالی طراحی شد که بر اساس متغیرهای ورودی به پیشگویی نوع سرطان پستان پرداخته است. سیستم طراحی‌شده با استفاده از مجموعه داده مربوط به بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین، موجود در انبار داده یادگیری ماشین دانشگاه ایروین، کالیفرنیا که شامل ۶۸۳ مورد بود، ارزیابی شد. داده‌های موجود در این مجموعه پیش‌پردازش شدند پس‌ازآن داده‌ها با روش خطی نرمال شدند. به‌منظور پیاده‌سازی شبکه از امکانات و توابع موجود در نرم‌افزار MATLAB بهره گرفته شد و از ۶۵٪ داده‌ها جهت مرحله آموزش شبکه و از ۳۵٪ باقیمانده جهت مرحله آزمون شبکه استفاده شد. از ۹ متغیر بالینی به‌عنوان ورودی شبکه استفاده شد. معیارهای حساسیت، اختصاصیت و صحت جهت ارزیابی در مرحله آزمون شبکه استفاده شد.

یافته‌ها: پس از شبیه‌سازی سیستم تصمیم‌یار با استفاده از شبکه عصبی احتمالی، پارامترهای حساسیت، اختصاصیت و صحت به کمک این سیستم به ترتیب معادل اعداد ۱، ۰/۹۸ و ۰/۹۹ به دست آمد.

نتیجه‌گیری: نتایج مطالعه نشان داد عملکرد شبکه عصبی احتمالی در تشخیص نوع سرطان پستان بهتر و قوی‌تر از سایر شبکه‌های عصبی مصنوعی بوده است. به‌علاوه شبکه پیاده‌سازی شده در این مقاله دارای سرعت بیشتر در فرایند آموزش و تعمیم‌پذیری بهتری نسبت به موارد مشابه بوده است.

واژه‌های کلیدی: اختصاصیت، حساسیت، سرطان پستان، سیستم تصمیم‌یار، شبکه عصبی احتمالی.
این مقاله مستخرج از پژوهشی است که با حمایت مالی دانشگاه پیام نور انجام‌شده است.

* نشانی نویسنده پاسخگو: فارس، شیراز، دانشگاه پیام نور، گروه مهندسی کامپیوتر و فناوری اطلاعات، آسیه خسروانیان.
نشانی الکترونیک: khosravanian.a@gmail.com

www.SID.ir

مقدمه

سرطان پستان رایج‌ترین شکل سرطان در زنان است. به گزارش مرکز تحقیقات سرطان دانشگاه علوم پزشکی شهید بهشتی در اسفند ماه سال ۱۳۹۲ شمسی حدود ۶۰ هزار نفر در ایران مبتلا به سرطان پستان هستند و سالانه حدود ۱۰ هزار مورد جدید هم به آمار مبتلایان به سرطان پستان در ایران اضافه می‌شود به همین دلیل کارشناسان می‌گویند که ایران بالاترین رشد ابتلا به سرطان پستان در جهان دارد. علاوه بر این، سن ابتلا به سرطان پستان در ایران ۵ تا ۱۰ سال پایین‌تر از میانگین جهانی است (۱). اگر چه شیوع بیماری سرطان پستان در دهه گذشته بسیار بالا بوده ولی میزان مرگ‌ومیر بر اثر این بیماری کاهش یافته است (۲). سرطان پستان مانند سایر سرطان‌ها با رشد و تکثیر سریع و کنترل نشده قسمتی از بافت سینه به وجود می‌آید که بر اساس نوع خطر آن به دو دسته خوش‌خیم و بدخیم طبقه‌بندی می‌شود (۳).

تومورهای خوش‌خیم رشد غیرطبیعی دارند ولی به‌ندرت باعث مرگ فرد می‌شوند با این‌حال تعدادی از توده‌های خوش‌خیم نیز می‌توانند خطر ابتلا به سرطان پستان را افزایش دهند. همچنین در برخی از زنان با سابقه بیوپسی از توده‌های خوش‌خیم سینه نیز، خطر سرطان پستان افزایش‌یافته است. از طرف دیگر تومورهای بدخیم جدی‌تر بوده و تشخیص زودهنگام آن‌ها شانس درمان موفقیت‌آمیز را بالابرده است (۴). بنابراین پیش‌بینی و تشخیص سریع این سرطان شانس درمان را بالا می‌برد بعلاوه در هزینه‌های بالای درمان برای بیماران مبتلا به آن صرفه‌جویی می‌شود. ریسک فاکتورهای ابتلا به سرطان پستان شامل سن بالا، سابقه خانوادگی، مصرف مشروبات الکلی و دخانیات، زمان اولین قاعدگی قبل از ۱۲ سالگی و شروع یائسگی بعد از ۵۵ سالگی، زمان بارداری یا عدم بارداری، اضافه وزن و چاقی بعد از یائسگی و زندگی بی‌تحرک است (۳).

از آنجاکه در تحقیقات علوم پزشکی مسئله سلامت انسان مطرح است، پیش‌بینی درست نتایج اهمیت بیشتری می‌یابد بنابراین باید از روش‌هایی استفاده شود که پیش‌بینی بر اساس آن‌ها دارای کمترین خطا و بیش‌ترین اطمینان باشد. ازجمله روش‌هایی که توجه بسیاری از محققین را به خود معطوف داشته است شبکه‌های عصبی مصنوعی است.

شبکه‌های عصبی مصنوعی، سیستم‌ها و روش‌های محاسباتی نوینی برای یادگیری ماشین، نمایش دانش و در انتها اعمال دانش به‌دست‌آمده در جهت پیش‌بینی پاسخ‌های خروجی از سیستم‌های پیچیده هستند. ایده اصلی این شبکه‌ها تا حدودی الهام گرفته از شیوه کارکرد سیستم عصبی زیستی، برای پردازش داده‌ها و اطلاعات به‌منظور یادگیری و ایجاد دانش است. عنصر کلیدی این ایده ایجاد ساختارهایی جدید برای سامانه پردازش اطلاعات است (۵).

شبکه‌های عصبی احتمالی (Probabilistic Neural Network: PNN) یکی از موفق‌ترین و کاربردی‌ترین نوع از شبکه‌های عصبی مصنوعی بوده که بر اساس نظر کارشناسان ابزارهای قدرتمندی برای شناخت و طبقه‌بندی الگوها با بیش‌ترین احتمال موفقیت‌اند (۶). این شبکه‌ها برای اولین بار در سال ۱۹۸۸ میلادی توسط Specht معرفی شدند و از مهم‌ترین روش‌های با ناظر بوده که به‌منظور شناخت و طبقه‌بندی الگوها مورد استفاده قرار می‌گیرند (۷).

شبکه‌های عصبی احتمالی الگوها را بر اساس استراتژی بیزی و تخمین زننده‌های غیر پارامتریک توابع چگالی احتمال طبقه‌بندی می‌کنند و ازجمله روش‌هایی هستند که مبتنی بر الگوهای آماری موجود در داده عمل می‌کنند. این‌گونه از شبکه‌ها از الگوریتم‌های یادگیری یک‌بار گذر از داده‌ها استفاده کرده به این معنا که شبکه تنها در یک دور آموزش می‌بینند و نیاز به تکرار دورهای آموزشی نیست این یکی از بزرگ‌ترین مزیت‌های شبکه عصبی احتمالی به‌حساب می‌آید، چراکه روش‌هایی که از رویکردهای ذهنی به‌منظور به دست آوردن الگوهای موجود در داده‌ها استفاده می‌نمایند معمولاً مستلزم ایجاد مقدار زیادی از تغییرات کوچک در پارامترهای شبکه به‌منظور اصلاح تدریجی بازه خود می‌باشند و این به معنای زمان یادگیری طولانی است (۸).

شبکه‌های عصبی احتمالی یک اشکال عمده نیز دارند و آن نیاز به ذخیره‌سازی تمامی نمونه‌های آموزشی به‌منظور طبقه‌بندی الگوهای جدید است که این خود باعث اشغال حجم بزرگی از حافظه می‌شود. از آنجایی‌که حافظه کامپیوترها خیلی ارزان و پربازده هستند امروزه میزان ذخیره اطلاعات خیلی جای نگرانی ندارد (۹).

Archive of SID

www.SID.ir

در سال‌های اخیر در مقالات بسیاری شبکه‌های عصبی احتمالی به‌عنوان ابزار پیش‌بینی و شناخت الگو موردتوجه قرارگرفته و نتایج مطلوبی نیز در مقایسه با سایر روش‌ها ارائه کرده است (۱۰).

مطالعات نشان می‌دهد روش‌های متفاوتی در تشخیص سرطان پستان به‌کار گرفته شده است از جمله شبکه‌های بیزی، شبکه‌های عصبی، درخت تصمیم‌گیری، ماشین بردار پشتیبان و روش‌های فازی (۱۱، ۱۲) که در این قسمت به چند نمونه آن اشاره می‌کنیم.

صالحی و همکاران به مقایسه مدل شبکه عصبی مصنوعی و رگرسیون کاکس در پیش‌بینی بقای بیماران مبتلا به سرطان پستان پرداختند که نتایج حاصل، نشان‌دهنده برتری نسبی قدرت تشخیص شبکه عصبی مصنوعی در پیش‌بینی بقای بیماران مبتلا به سرطان پستان بود (۱۳). سدهی و همکاران به تحلیل بقای بیماران مبتلا به سرطان پستان در استان اردبیل با استفاده از شبکه عصبی مصنوعی و رگرسیون کاکس پرداختند. نتایج این مطالعه نشان داد کارایی مدل‌های شبکه عصبی مصنوعی برای پیش‌بینی وضعیت بقا بیماران مبتلا به سرطان پستان در مقایسه با رگرسیون کاکس وابسته به نوع معماری و الگوریتم آموزش شبکه متفاوت بوده و چنانچه معماری مناسبی برای مدل شبکه عصبی مصنوعی انتخاب گردد، این مدل نسبت به مدل رگرسیون کاکس کارایی بیشتری برای پیش‌بینی وضعیت بقا دارد (۱۴). دلن و همکاران از شبکه‌های عصبی مصنوعی، درخت تصمیم‌گیری و رگرسیون لجستیک برای توسعه مدل‌های پیش‌بینی سرطان پستان با تجزیه‌وتحلیل پایگاه‌های بزرگ داده که از پایگاه داده مشهور ویسکانسین گردآوری‌شده بود، بهره جستند. نتایج تحقیقات آن‌ها نشان داد که الگوریتم درخت تصمیم برای استخراج دانش از داده‌های موجود، مقدم بر سایر روش‌ها بود و نتایج به‌دست‌آمده از تحقیق، نزدیک به واقعیت بود (۱۵). پندهارکر و همکاران برای بررسی الگوهای موجود در سرطان پستان از چندین روش داده‌کاوی استفاده کردند. نتیجه مطالعه آن‌ها نشان داد داده‌کاوی می‌تواند به نحو مطلوبی در شناسایی الگوها در تشخیص سرطان پستان به کار رود (۱۶). لاندین و همکارانش از مدل‌های شبکه عصبی مصنوعی و رگرسیون لجستیک برای پیش‌بینی ۵، ۱۰ و ۱۵ ساله بقای بیماران مبتلا به سرطان پستان استفاده کردند و ۹۵۱ بیمار مبتلا به سرطان پستان را مورد مطالعه قرار دادند و به این نتیجه رسیدند که درختان طبقه‌بندی و همچنین رگرسیون لجستیک برای تفسیر بالینی بسیار آسان‌تر است (۱۷). همچنین یی و فویانگ از ماشین بردار پشتیبان (SVM) برای کشف الگوهای تشخیص سرطان پستان استفاده کردند. نتایج مطالعه آن‌ها نشان داد که SVM برای تشخیص الگوهای سرطان پستان روش مناسبی بود و نتایج به‌دست‌آمده با شواهد موجود و واقعی مطابقت داشت (۱۸).

فوگارتی و همکاران الگوریتم ژنتیک را برای تشخیص سرطان پستان بر روی مجموعه داده ویسکانسین به کاربردند و پس از مقایسه خروجی با خروجی واقعی به دقت ۹۶/۳۲٪ رسیدند (۱۹). ایران پور و همکاران شبکه‌های RBF و SVM را برای تشخیص سرطان پستان موردبحث قراردادند و به دقت ۹۸/۱٪ رسیدند (۲۰). سلطانی سروستانی و همکاران مقایسه‌ای بین شبکه‌های عصبی مصنوعی MLP، SOM، PNN و RBF بر دو مجموعه داده ویسکانسین و بیمارستان نمازی شیراز در تشخیص سرطان پستان انجام دادند نتایج این تحقیق حاکی از برتری شبکه‌های PNN و RBF در تشخیص این بیماری بود (۲۱).

هدف مطالعه حاضر، طراحی، پیاده‌سازی و ارزیابی یک سیستم تصمیم یار جهت کمک به تشخیص پزشک در تعیین نوع توده‌های سرطان پستان با استفاده از شبکه عصبی احتمالی است.

مواد و روش‌ها

در این مطالعه یک شبکه عصبی احتمالی طراحی شد که بر اساس متغیرهای ورودی به پیشگویی نوع سرطان پستان پرداخته است. سیستم طراحی‌شده با استفاده از مجموعه داده مربوط به بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین، موجود در انبار داده یادگیری ماشین دانشگاه ایروین، کالیفرنیا (UCI: University of California, Irvine) در دسترس در (۲۲) که شامل ۶۹۹ مورد بود، ارزیابی شد. به‌منظور پیاده‌سازی شبکه از امکانات و توابع موجود در نرم‌افزار MATLAB بهره گرفته شد. همان‌طور که در جدول ۱ نشان داده‌شده است مطابق با مطالعات پیشین از ۹ متغیر بالینی به‌عنوان ورودی شبکه استفاده شد. متغیرهای ورودی شامل

www.SID.ir

فاکتورهای ضخامت انبوه، یکنواختی اندازه سلول، یکنواختی شکل سلول، چسبندگی لبه‌ها، حجم سلول بافت اپیتلیال، هسته‌های عریان، کروماتین بلاند، هسته عادی و تقسیم هسته سلول به دو قسمت بود.

جدول ۱: متغیرهای ورودی در تشخیص سرطان پستان

متغیرهای بالینی	محدوده	ورودی شبکه عصبی مصنوعی
ضخامت انبوه	۱-۱۰	۱۰/عدد
یکنواختی اندازه سلول	۱-۱۰	۱۰/عدد
یکنواختی شکل سلول	۱-۱۰	۱۰/عدد
چسبندگی لبه‌ها	۱-۱۰	۱۰/عدد
حجم سلول بافت اپیتلیال	۱-۱۰	۱۰/عدد
هسته‌های عریان	۱-۱۰	۱۰/عدد
کروماتین بلاند	۱-۱۰	۱۰/عدد
هسته عادی	۱-۱۰	۱۰/عدد
تقسیم هسته سلول به دو قسمت	۱-۱۰	۱۰/عدد

داده‌های موجود در این مجموعه به سه دلیل پیش‌پردازش شدند:

۱- در این مجموعه داده، مقدار بعضی داده‌ها وارد نشده بود که به اصطلاح داده‌های گم شده بودند، هدف از این پیش‌پردازش این بود که داده‌های گم شده حذف شوند که در روند شبیه‌سازی خطا را کم کنند. سطرهایی که شامل داده‌های گم‌شده بودند حذف شدند که پس از این مرحله ۶۸۳ سطر باقی ماند.

۲- از آن جا که شماره شناسایی بیماران جزء فاکتورهای تشخیصی نوع سرطان پستان نبود در پیش‌پردازش داده‌ها حذف شدند.

۳- در پایان داده‌ها با روش خطی نرمال شدند به‌گونه‌ای که مقادیر عددی آن‌ها بین صفر و یک قرار گیرد که از نظر فنی قابل اعمال به شبکه عصبی احتمالی شود.

به‌طورکلی برای بررسی میزان موفقیت و کارایی سیستم‌های دسته‌بندی و تشخیص بیماری‌ها، از ماتریس کانفیوژن استفاده می‌شود. تحلیل‌های ماتریس کانفیوژن در دسته‌بندی و تشخیص بیماران منجر به ۴ حالت مثبت حقیقی (True Positive یا TP)، منفی حقیقی (True Negative یا TN)، مثبت کاذب (False Positive یا FP) و منفی کاذب (False Negative یا FN) می‌شود. از نتایج ماتریس کانفیوژن سه شاخص حساسیت (دقت سیستم در تشخیص نوع بدخیم) و اختصاصیت (دقت سیستم در تشخیص نوع خوش‌خیم) و صحت (نسبت تمام مواردی که به‌صورت صحیح طبقه‌بندی شدند) به دست می‌آید که برای تحلیل عملکرد سیستم‌های طبقه‌بندی استفاده می‌شود.

$$\text{حساسیت} = \frac{TP}{TP+FN} \quad (۱)$$

$$\text{اختصاصیت} = \frac{TN}{FP+TN} \quad (۲)$$

$$\text{صحت} = \frac{TP+TN}{TP+TN+FP+FN} \quad (۳)$$

مرحله آموزش شبکه عصبی:

در این مرحله از ۶۵ درصد داده‌ها (۴۴۴ نمونه) جهت آموزش شبکه عصبی مصنوعی استفاده شده است. برای پیاده‌سازی شبکه عصبی احتمالی در نرم‌افزار MATLAB از یک ماتریس ورودی شامل ۹ سطر (۹ متغیر بالینی) و ۴۴۴ ستون و ماتریس دیگری با ۲ سطر (۲ کلاس خوش‌خیم و بدخیم) و ۴۴۴ ستون به‌عنوان ماتریس هدف استفاده شد. داده‌های ورودی به شبکه با روش خطی نرمال شدند به‌گونه‌ای که داده‌ها بین صفر و یک قرار گیرد و سپس به شبکه عصبی احتمالی وارد شده‌اند. ماتریس هدف نشان‌دهنده دو کلاس خوش‌خیم و بدخیم بود. درصورتی‌که نوع سرطان متعلق به کلاس مربوطه بود سطر مربوط به آن با عدد یک و سطر دیگر با عدد صفر پر شد.

در شبکه عصبی احتمالی طراحی‌شده، تنها از یک دور برای آموزش شبکه استفاده شد که این از مزایای مهم شبکه‌های عصبی احتمالی نسبت به دیگر شبکه‌هاست.

مرحله آزمون شبکه عصبی:

در این مرحله ۳۵٪ داده‌ها (۲۳۹ نمونه) که در مرحله آموزش استفاده‌نشده بودند، به‌صورت بردار به شبکه عصبی مصنوعی پیاده‌سازی شده در نرم‌افزار اعمال گردید.

Archive of SID
www.SID.ir

یافته‌ها

خروجی شبکه عصبی احتمالی طراحی‌شده در مرحله آزمون و مقایسه آن با نتایج واقعی جمع آوری شده از بیماران مورد مطالعه در جدول ۲ نشان داده شده است.

جدول ۲ : پیش‌بینی نوع سرطان پستان در مرحله آزمون شبکه عصبی احتمالی

دسته‌بندی	پیش‌بینی شبکه عصبی احتمالی	نتایج واقعی
کلاس خوش‌خیم	۱۸۲	۱۸۴
کلاس بدخیم	۵۷	۵۵

همان‌طور که مشاهده می‌شود از ۲۳۹ نمونه که در مرحله آزمون مورداستفاده قرار گرفت ۱۸۴ نفر مبتلا به سرطان خوش‌خیم سینه و ۵۵ نفر دچار نوع بدخیم آن بودند، شبکه عصبی احتمالی ۱۸۲ نفر را مبتلا به سرطان خوش‌خیم و ۵۷ نفر را مبتلا به سرطان بدخیم تشخیص داده است که به این معناست که شبکه دو نوع خوش‌خیم را به‌اشتباه بدخیم تشخیص داده است.

برای مجموعه داده‌های به‌کاررفته در این مقاله حساسیت و اختصاصیت و صحت با استفاده از شبکه عصبی احتمالی به ترتیب اعداد ۱ و ۰/۹۸ و ۰/۹۹ به‌دست‌آمده است.

بحث

مطالعه فعلی باهدف طراحی، پیاده‌سازی و ارزیابی یک سیستم تصمیم یار جهت کمک به تشخیص پزشک در تعیین نوع توده‌های سرطان پستان با استفاده از شبکه عصبی احتمالی انجام شد.

سیستم تصمیم‌یار طراحی‌شده در این مقاله در تشخیص توده‌های خوش‌خیم و بدخیم موفق بوده است و دسته‌بندی را با دقت مناسبی انجام داده است. حسن این طراحی انتخاب مناسب شبکه عصبی مصنوعی احتمالی بود که در این شبکه فرایند آموزش شامل یک مرحله است و هیچ‌گونه تکراری برای تعدیل وزن‌ها نیاز نیست. استفاده از شبکه عصبی احتمالی در تشخیص نوع سرطان پستان روش پیشنهادی در این مقاله بود که نتایج حاصل از آن حاکی از سرعت بیشتر و تعمیم‌پذیری بهتر، نسبت به دیگر موارد پیاده‌سازی شده در این موضوع بود. نتایج شبیه‌سازی نشان داد سیستم تصمیم‌یار طراحی‌شده در این مقاله بر روی مجموعه داده بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین به دقت (صحت) ۹۹٪ رسیده است که بالاتر از تحقیقات مشابه بر روی این مجموعه داده بوده است. به نظر می‌آید مقایسه نتایج مطالعه با مطالعاتی که بر سایر مجموعه‌های داده انجام شده است صحیح نباشد زیرا که داده‌های مورداستفاده متفاوت است و تکنیک های مختلفی با توجه به نوع داده انتخاب شده است ازاین‌رو در مقایسه نتایج این مطالعه با مطالعات دیگر ابتدا به مطالعاتی که مجموعه داده یکسان با این پژوهش (مجموعه داده سرطان پستان بیمارستان ویسکانسین) داشته و پیاده‌سازی با شبکه‌های عصبی مصنوعی متفاوت صورت گرفته اشاره شده است.

فوگارتی و همکاران الگوریتم ژنتیک را برای تشخیص سرطان پستان بر روی مجموعه داده ویسکانسین به کاربردند و پس از مقایسه خروجی با خروجی واقعی به دقت ۹۶/۳۲٪ رسیدند (۱۹). ایران پور و همکاران شبکه‌های RBF و SVM را برای تشخیص سرطان پستان مورد بحث قراردادند و به دقت ۹۸/۱٪ رسیدند (۲۰).

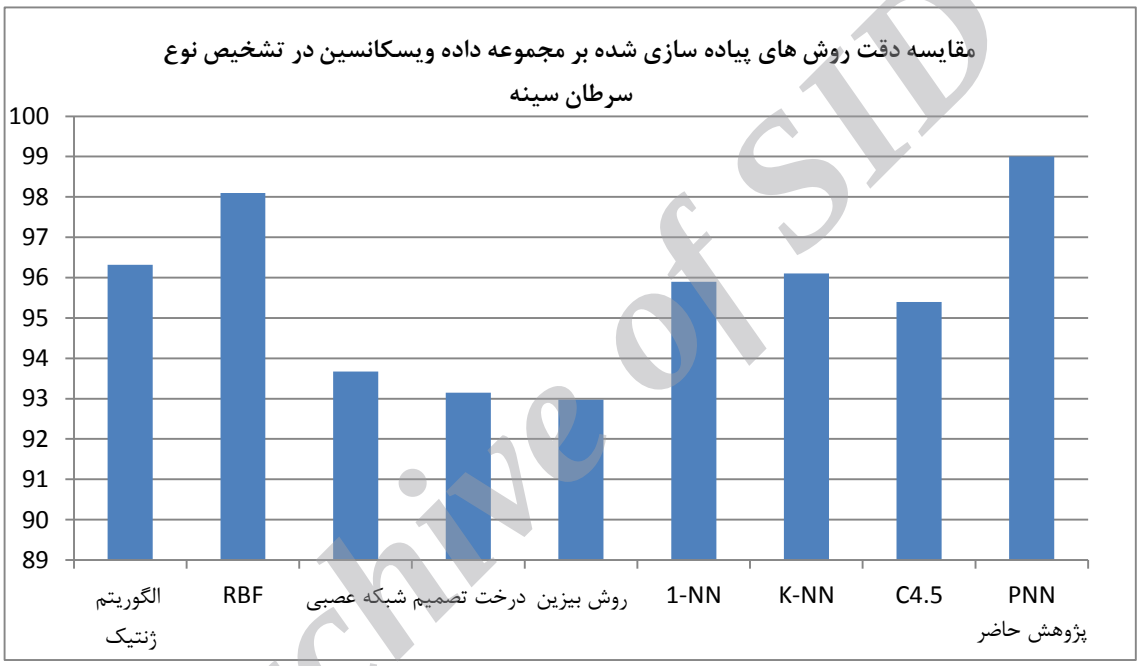
در مطالعه‌ای دیگر نتایج پیاده‌سازی بر روی مجموعه داده بیمارستان ویسکانسین در روش بیزین به دقت ۹۲/۶۱٪، شبکه عصبی RBF به دقت ۹۳/۶۷٪ و درخت تصمیم به دقت ۹۲/۹۷٪ دست‌یافته است (۲۳). سالاما و همکاران با استفاده از روش بیزین ساده و درخت تصمیم بر روی مجموعه داده بیمارستان ویسکانسین به ترتیب به دقت ۹۲/۹۷٪ و ۹۳/۱۵٪ رسیدند (۲۴). در مقایسه با نتایج به‌دست‌آمده در مطالعات ذکر شده پیشین، در این پژوهش نیز از همان مجموعه داده استفاده شد و با انتخاب شبکه عصبی احتمالی که روشی متفاوت با مقالات ذکرشده بود طراحی و ارزیابی صورت گرفت که نتیجه آن رسیدن به دقت ۹۹٪ بوده است. جدول ۳ مقایسه بین دقت به‌دست‌آمده در منابع مذکور و پژوهش حاضر را نشان می‌دهد.

سلطانی سروستانی و همکاران روش‌های مختلفی جهت تشخیص خوش‌خیم یا بدخیم بودن سرطان پستان با استفاده از شبکه‌های عصبی مختلف بر روی دو دسته داده ارائه کردند و متوسط مربع خطا در هر شبکه را باهم مقایسه کردند که از میان شبکه‌های عصبی چند لایه،

www.SID.ir

جدول ۳: مقایسه صحت داده‌ها در پیاده‌سازی‌های مختلف در تشخیص سرطان پستان

نویسندگان	پایگاه داده مورد استفاده در سرطان سینه	روش پیاده‌سازی	دقت در مرحله آزمون	سال
فوگارتی و همکاران (۱۹)	مجموعه داده بیمارستان ویسکانسین	الگوریتم ژنتیک	۹۶/۳۲	۲۰۰۱
ایران پور و همکاران (۲۰)	مجموعه داده بیمارستان ویسکانسین	RBF و SVM	۹۸/۱	۲۰۰۷
آرانا و همکاران (۲۳)	مجموعه داده بیمارستان ویسکانسین	شبکه عصبی RBF	۹۳/۶۷	۲۰۱۱
آرانا و همکاران (۲۳)	مجموعه داده بیمارستان ویسکانسین	درخت تصمیم	۹۲/۹۷	۲۰۱۱
آرانا و همکاران (۲۳)	مجموعه داده بیمارستان ویسکانسین	روش بیزین	۹۲/۶۱	۲۰۱۱
سالاما و همکاران (۲۴)	مجموعه داده بیمارستان ویسکانسین	روش بیزین	۹۲/۹۷	۲۰۱۲
سالاما و همکاران (۲۴)	مجموعه داده بیمارستان ویسکانسین	درخت تصمیم	۹۳/۱۵	۲۰۱۲
پژوهش حاضر	مجموعه داده بیمارستان ویسکانسین	شبکه عصبی احتمالی	۹۹	۲۰۱۵



شکل ۱: مقایسه دقت روش‌های پیاده‌سازی شده بر مجموعه داده ویسکانسین در تشخیص نوع سرطان پستان

رقابتی و پایه شعاعی، شبکه عصبی پایه شعاعی به کمترین متوسط خطا و در نتیجه دقت بهتری دست یافت (۲۱). در پژوهش دیگری مدحت محمد احمد و همکاران از طبقه‌بندی کننده SVM استفاده کردند و به این نتیجه رسیدند که نمودار ROC در این روش دارای سطح زیر منحنی بیشتری است و این نشان می‌دهد که به دقت خوبی دست یافتند (۲۵). به نقل از (۲۶) استفاده از الگوریتم‌های 1-NN و K-NN و C4.5 بر این مجموعه داده به ترتیب به دقت ۹۵/۹، ۹۶/۱ و ۹۵/۴ رسیده است. نتایج مطالعات ذکرشده محققان بر روی این مجموعه داده در مقایسه با سیستم تصمیم‌یار طراحی‌شده در این مقاله نشان می‌دهد که شبکه عصبی احتمالی در این پیاده‌سازی و در مقطع زمانی فعلی به دقت بهتری رسیده است. مقایسه نشان می‌دهد سیستم موردمطالعه قابلیت رقابت با سیستم‌های قبلی معرفی‌شده را دارد. شکل ۱ دقت روش‌های پیاده‌سازی شده تاکنون بر روی مجموعه داده ویسکانسین در تشخیص نوع سرطان پستان را نشان می‌دهد.

از محدودیت‌های این پژوهش می‌توان به ۱۹ رکورد حذف شده از مجموعه داده اشاره کرد. داده‌های گم شده تخمین زده نشدند و جایگزین نشدند که می‌تواند بر نتایج به‌دست‌آمده از این پژوهش مؤثر باشد. گرچه در دسترس بودن سیستم کامپیوتری آموزش‌دیده با داده‌های مورد نظر هم از نظر تکنیکی، خود محدودیت استفاده است. از دیگر

Archive of SID
www.SID.ir

محدودیت‌ها به‌روز نبودن مجموعه داده است که با توجه به اینکه هر سال آمار جدیدی از مبتلایان به این بیماری منتشر می‌شود بهتر است با داده‌های واقعی و به‌روز که تعداد کافی از بیماران را شامل می‌شود مدلسازی انجام شود و نتایج ارزیابی شود. در صورتی که در تهیه مجموعه داده از نظرات پزشک متخصص استفاده شود و تمام موارد خاص مربوط به سرطان پستان در نظر گرفته شود پایگاه جامعی خواهیم داشت که مدلسازی را به دقتی نزدیک به واقعیت می‌رساند. همچنین این پژوهش بیشتر پیاده‌سازی سیستمی برای کمک به پزشک است لذا جنبه پزشکی موضوع موردمطالعه، مد نظر نویسندگان قرار نگرفته و در این زمینه به یافته‌های علمی پزشکان استناد شده است. در واقع هدف، پیاده‌سازی یک سیستم تصمیم یار پزشک مبتنی بر یک شبکه عصبی جدید و آموزش‌دیده با مجموعه داده مورد نظر و مقایسه نتایج با سایر پژوهش‌ها بوده است. در راستای تکمیل این پژوهش در گام بعد می‌توان پس از ارزیابی این سیستم، با مشورت پزشکان متخصص ریسک فاکتورها را با دقت بیشتری انتخاب کرده و همچنین برای بومی کردن این سیستم، شبکه را با مجموعه داده‌های بیمارستان‌های ایران آموزش داد و به پیش‌بینی این بیماری در کشور ایران پرداخت.

نتیجه‌گیری

در این مقاله از شبکه عصبی احتمالی برای طبقه‌بندی نوع سرطان پستان به دو دسته خوش‌خیم و بدخیم استفاده شد. ابتدا داده‌های مورد استفاده پیش‌پردازش شدند سپس شبکه با نمونه‌های آموزشی، آموزش داده شد و در انتها با نمونه‌های آزمون، آزمون شد. از دلایل بالا بودن حساسیت و اختصاصیت در مقاله حاضر می‌توان به پیش پردازش داده‌های ورودی همانطور که پیش‌تر شرح داده شد و انتخاب مناسب شبکه عصبی برای این منظور اشاره کرد.

در واقع شبکه پیاده‌سازی شده در این مقاله به خاطر سرعت زیاد و تعمیم‌پذیری خوب آن نسبت به دیگر شبکه‌های عصبی مصنوعی برتر است. پیشنهاد می‌شود نرم‌افزاری بر اساس این شبیه‌سازی طراحی شود و از نتایج حاصل از این پژوهش در آموزش پزشکان استفاده شود.

References

1. Akbari MS. Ten thousand Iranian women each year are diagnosed with breast cancer. 2014. Cancer Research Center. Available at: http://crc.sbmu.ac.ir/?fkeyid=&siteid=148&pageid=11062. 2014.
2. Haddadnia J. A Novel Smart System for Mammographic Image Classification Based on Breast Density. Iranian Journal of Breast Disease 2013; 6 (1) :15-22 :URL http://www.ijbd.ir/browse.php?a_code=A-10-222-10&slc_lang=en&sid=1 .
3. National Center for Chronic Disease Prevention and Health Promotion, Breast Cancer Division, Breast Cancer and You fact sheet. Dec 2010. Available online at: www.cdc.gov/cancer/breast/pdf/BreastCancerFS_Dec2010.pdf.
4. American Cancer Society (2006): URL http://www.cancer.org.
5. Ghasem Ahmad L. Review top 7 Algorithms in Data Mining for Prediction Survivability, Diagnosis and Recurrence of Breast Cancer. Iranian Journal of Breast Disease. 2013; 6 (1): 52-61. URL http://www.ijbd.ir/browse.php?a_code=A-10-222-14&slc_lang=fa&sid=1 .
6. Wasserman P. Advanced Methods in Neural Computing. New York: Van Nostrand Reinhold 1993.
7. Specht D. Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. IEEE International Conference on Neural Networks 1988; 1:525-32.
8. Ghavam-Zadeh M. Forecasting in contracts organized markets. Master of Science Thesis, Tehran University, Electric and Electronic Engineering Department 1998.
9. Cho G, Kim T, Seo Y, Chan M. Integrated machining error compensation method using OMM data and modified PNN algorithm. International Journal of Machine Tools & Manufacture 2006; 46(12-13): 1417-27.
10. Kalatzis I, Liappas I. Design and Implementation of a Multi-PNN Structure for Discriminating One-Month Abstinent Heroin Addicts From Healthy Controls

Archive of SID
www.SID.ir

Using the P600 Component of ERP Signals. Pattern Recognition Letters 2005; 26: 1691– 700.

11. Abbass HA. An evolutionary artificial neural networks approach for breast cancer diagnosis. Artificial Intelligence in Medicine 2002; 25(3): 265-81.

12. Jiang L, Wang D, Cai Z, Yan X. Survey of improving naive Bayes for classification. Advanced Data Mining and Applications 2007: 134-45.

13. Salehi M, gohari M, vahabi N, Zayeri F, yahyazadeh S, kafashian M. Comparison of Artificial Neural Network and Cox Regression Models in Survival Prediction of Breast Cancer Patients. journal of ilam university of medical sciences 2013; 21 (2):120-8. URL http://sjimu.medilam.ac.ir/browse.php?a_code=A-10-220-127&slc_lang=fa&sid=1.

14. Sedehi M, Amani F, Momeni Dehaghi F. Analysis of survival data of patient with breast cancer using artificial neural network and cox regression models. Journal of zabol university of medical sciences and health services 2014; 5 (4): 24-32. URL http://www.zbmu.ac.ir/journal/browse.php?a_code=A-10-285-1&slc_lang=fa&sid=1.

15. Ghasem Ahmad L. Using Data Mining Techniques for Prediction Breast Cancer Recurrence. Iranian Journal of Breast Disease 2013; 5(4):2 3-34. URL http://www.ijbd.ir/browse.php?a_code=A-10-222-2&slc_lang=fa&sid=1.

16. Pendharkar PC, Rodger JA, Yaverbaum GJ, Herman N, Benner M. Associations statistical, mathematical and neural approaches for mining breast cancer patterns. Expert Systems with Applications 1999; 17: 223–32.

17. Lundin M, Lundin J, Burke HB,Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. Oncology 1999; 57(4): 281-6.

18. Yi W, Fuyong W. Breast cancer diagnosis via support vector machines. In Proc. the Twenty.

19. Werner JC, Fogarty TC. Genetic Programming Applied to Severe Diseases Diagnosis. In Proceedings Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP), 2001.

20. Iranpour M, Almassi S, Analoui M, Breast Cancer Detection from FNA using SVM and RBF Classifier. In 1st Joint Congress on Fuzzy and Intelligent Systems 2007.

21. Sarvestan Soltani A, Safavi AA, Parandeh MN, Salehi M. Predicting Breast Cancer Survivability using data mining techniques. Software Technology and Engineering (ICSTE). 2nd International Conference 2010; 2: 227-31.

22. UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original).

23. Aruna S, Rajagopalan DS, Nandakishore LV. Knowledge based analysis of various statistical tools in detecting breast cancer. Comput Sci Inform Tech 2011; 2: 37-45.

24. Salama GI, Abdelhalim MB, Zeid MAE. Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. Int J Comput Sci Inform Tech 2012; 1(1): 2277- 0764.

25. Mohamed Medhat A, Muhamed FW. Using data mining for assessing diagnosis of breast cnacer. International multiconfrence on computer science and information Technology 2010; 11-17.

26. Mahmoodi SA, Mahmoodi MS, Abbasniya VS, Mahmoodi SM. Assessment of Classification Algorithms in the Diagnosis of Diabetes and Breast Cancer. Iranian Journal of Medical Informatics 2012; 2(2):19-22.

Archive of SID

www.SID.ir