

پیش بینی بقای سرطان سینه به کمک شبکه یادگیری عمیق و داده‌های چند وجهی

تاریخ دریافت: ۱۴۰۳/۰۱/۰۹ - تاریخ پذیرش: ۱۴۰۳/۰۶/۱۹

خلاصه

مقدمه

سرطان سینه به‌عنوان یکی از سرطان‌های رایج و مهم، نقش بسزایی در افزایش نرخ مرگ و میر در زنان دارد. در حال حاضر، داده‌های چندوجهی مرتبط با سرطان، از جمله جزئیات ژنومی، تصاویر ماموگرافی و اطلاعات بالینی، در دسترس قرار گرفته است که این امر باعث شده تا توجه بیشتری به توسعه مدل‌های پیشرفته یادگیری عمیق برای پیش‌بینی بقای سرطان سینه شود. در این تحقیق، یک مدل یادگیری عمیق مبتنی بر شبکه‌های عصبی پیچشی با استفاده از داده‌های چندوجهی برای پیش‌بینی بقای سرطان سینه پیشنهاد می‌شود که ویژگی‌های مهمی را برای افزایش دقت این پیش‌بینی تولید می‌کند. روش پیشنهادی دارای دو فاز است؛ در فاز اول از یک شبکه عصبی پیچشی برای استخراج ویژگی‌های موثر استفاده می‌شود و در فاز دوم، این ویژگی‌ها، برای انجام فرآیند پیش‌بینی بقای سرطان سینه به کار گرفته می‌شوند. نتایج آزمایشات نشان داد که در حالتی که از داده‌های بالینی استفاده گردید دقت روش پیشنهادی ۹۸/۳۴ حاصل گردید و در حالت بدون استفاده از این داده‌ها دقت ۹۷/۲۱ حاصل گردید که حاکی از این است مدل پیشنهادی این تحقیق نسبت به مدل‌های مقایسه‌شده، بهبود قابل توجهی داشته است. به‌علاوه، نتایج نشان می‌دهند که استفاده از داده‌های بالینی در ارتقاء دقت و عملکرد مدل پیشنهادی مؤثر می‌باشد.

کلمات کلیدی

سرطان سینه، ماموگرافی، پیش‌بینی بقای سرطان سینه، یادگیری عمیق، شبکه عصبی پیچشی.
بی‌نوشت: این مطالعه فاقد تضاد منافع می‌باشد.

فاطمه جلالی^۱

محمد مهدی حسینی^{۲*}

^۱ کارشناسی ارشد، گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه عالی شاهرود (غیرانتفاعی)، شاهرود، ایران.

^۲ استادیار، گروه کامپیوتر، دانشکده فنی و مهندسی، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

Email:

Hosseini_mmm@shahroodut.ac.ir

مقدمه

سرطان ممکن است تقریباً از هر قسمتی از بدن انسان که سلول‌ها شروع به رشد غیرقابل کنترل می‌کنند، ایجاد شود. نتایج تحقیقات نشان داده است که مرگ و میر ناشی از سرطان هر سال افزایش می‌یابد و به عنوان اصلی‌ترین بیماری در سطح جهانی در نظر گرفته می‌شود (۱). بنابراین تشخیص و درمان زودهنگام ممکن است در جلوگیری از پیشرفت سرطان به سطح پیشرفته سرطان مفید باشد. سرطان سینه^۱ یکی از بیماری‌های مهمی است که به بالاترین میزان مرگ و میر در میان زنان تبدیل شده است (۲). ساختار سینه زنان عمدتاً از مجاری شیر، لوبول‌ها و بافت چربی تشکیل شده است. سرطان سینه ممکن است در مجاری که شیر را به نوک پستان می‌برند یا در غدد لوبول (بخشی از پستان که شیر مادر را تولید می‌کند) شروع شود. در سطح جهانی با توجه به اینکه ۴۰ تا ۷۵ درصد از زیرگروه‌های مجرای همه موارد گزارش شده را تشکیل می‌دهند، اکثر سرطان‌های سینه از زیرگروه‌های مجاری و لوبولار هستند (۱). سرطان سینه به‌عنوان یکی از نوع‌های خطرناک و تهاجمی‌ترین سرطان‌ها، با توانایی ظاهر شدن به صورت توده‌های سینه، امروزه مورد توجه بسیاری از محققان گرفته و سبب گردیده تا بسیاری از تحقیقات در این زمینه مورد بررسی قرار گیرند. تأثیر تغییرات ژنتیکی بر روی این نوع سرطان، محققان را به فکر بیان ژن و تنوع تعداد کپی آنها در ارتباط با جزئیات بالینی سرطان پستان انداخته است (۲). رشد سرطان پستان، شامل فرآیندهای میانی است که از پروليفراسیون مجرای شروع شده و با تکامل به کارسینوم درجا، کارسینوم تهاجمی و در نهایت، بیماری متاستاتیک پیش می‌رود (۳). دو نوع مهم سرطان سینه تهاجمی و غیرتهاجمی هستند که تفاوت‌های مهمی در رفتار و گسترش خود دارند.

سرطان تهاجمی پستان از طریق گسترش در بافت‌های مجاور بدن پیش می‌رود در حالی که سرطان غیرتهاجمی محدود به غدد لوبول یا مجاری شیر در سینه می‌شود. با توجه به منشأ

این نوع سرطان از لوبول‌ها، به عنوان یکی از شکل‌های تهدید کننده سرطانی در نظر گرفته می‌شود که علل اصلی مرگ و میرهای مرتبط با سرطان در جهان است (۴). تخصصی‌های پزشکی در زمینه پیش‌بینی و درمان سرطان پستان تهاجمی با چالش‌های بسیاری مواجه هستند، زیرا درک اطلاعات مرتبط از منابع مختلف برای انسان یک مسئله پیچیده است (۵). تحقیقات نشان داده با گذر زمان به علت عدم توجه به این بیماری در مراحل اولیه، مبتلایان با مشکلات حادثتری مواجه می‌شوند و عدم وجود درمان قطعی برای آن، روند پیشرفت آن قابل کنترل نخواهد بود. با توجه به اینکه بسیاری از افراد مبتلا به این بیماری در مراحل اولیه توانسته‌اند وجود بقایا بعد از درمان سرطان سینه را تشخیص و از پیشرفت آن جلوگیری نمایند. بنابراین زمان در شناسایی و کنترل این بیماری بسیار مهم می‌باشد. اگرچه راهکارهای ارائه شده در تحقیقات اخیر پیشرفت‌های مناسبی در این حوزه داشته‌اند، اما بررسی بیشتر آنها نشان می‌دهد که بکارگیری هر یک از آن روش‌ها نیازمند مهیاسازی برخی شرایط خاص است که در وضعیت واقعی، دستیابی به این شرایط امکان‌پذیر نیست یا اینکه با افزایش میزان داده‌ها کارایی خود را از دست می‌دهند و سرعت یا دقت پیش‌بینی آنها کاهش می‌یابد. به همین دلیل، متخصصان حوزه هوش مصنوعی به دنبال ارائه راهکارهایی بر مبنای روش‌های یادگیری ماشین و یادگیری عمیق برای تشخیص و پیش‌بینی این بیماری هستند و مطالعات زیادی نیز در این حوزه انجام شده است که نشان می‌دهد، بکارگیری روش‌های سنتی یادگیری ماشین در حوزه پیش‌بینی سرطان سینه کارایی و دقت مناسبی ندارند و نیاز است تا از روش‌های مناسب‌تر در این حوزه استفاده گردد (۵-۷). از این رو، نیاز به ایجاد یک راهکار دقیق و کارا برای پیش‌بینی بقای سرطان سینه با استفاده از یادگیری عمیق پیشرفته چندوجهی ضروری و مهم به نظر می‌رسد. این پژوهش در تلاش برای ارائه یک مدل پیش‌بینی هوشمند برای پیش‌بینی شانس بقا است که می‌تواند

¹ Breast Cancer

سرطان است که بر نحوه طبقه‌بندی و اطلاع‌رسانی سرطان نظارت می‌کنند. این امر برای اطمینان از این است که همه پزشکان و مراکز درمانی سرطان را به شیوه‌ای یکسان توصیف کنند تا نتایج درمان همه افراد قابل مقایسه و درک باشد. در گذشته، تعداد مرحله تنها بر اساس سه حساسیت بالینی T، N و M محاسبه می‌شد. دسته T (اندازه) تومور اصلی (اولیه) شامل موارد زیر است (۹):

- TX: تومور قابل ارزیابی نیست.
- T0: هیچ شواهدی از تومور اولیه وجود ندارد.
- T: سرطان در محل است (تومور شروع به رشد به بافت سالم پستان نکرده است).
- T1، T2، T3، T4: این اعداد بر اساس اندازه تومور و میزان رشد آن به بافت سینه همسایه است. هرچه عدد T بالاتر باشد، تومور بزرگتر و/یا بیشتر در بافت پستان رشد کرده است.

دسته N (درگیری غدد لنفاوی) توضیح می‌دهد که آیا سرطان به غدد لنفاوی مجاور رسیده است یا خیر (۸، ۹):

- NX: غدد لنفاوی مجاور را نمی‌توان ارزیابی کرد، به عنوان مثال، اگر قبلاً برداشته شده باشند.
- N0: غدد لنفاوی مجاور حاوی سرطان نیستند.
- N1، N2، N3: این اعداد بر اساس تعداد غدد لنفاوی درگیر و میزان سرطان در آنها می‌باشد. هرچه عدد N بیشتر باشد، میزان درگیری غدد لنفاوی بیشتر است.

دسته M (متاستاز) نشان می‌دهد که آیا شواهدی وجود دارد که سرطان به سایر قسمت‌های بدن سفر کرده است یا خیر (۸، ۹):

- MX: متاستاز قابل ارزیابی نیست.
- M0: متاستاز دور وجود ندارد.
- M1: متاستاز دور دست وجود دارد.

تیم مراقبت‌های بهداشتی درمان را بر اساس مرحله بندی سرطان سینه توصیه می‌کنند. از طرفی، درمان‌های خاص نیز به عوامل فردی زیادی بستگی دارد، بنابراین می‌توان

به بیماران با یک برنامه درمانی شخصی و خاص برای درمان سرطان کمک کند. همچنین می‌تواند به پزشکان اجازه دهد تا تصمیمات درمانی مناسبی را برای بیماران اتخاذ کنند. مدل ارائه شده در این پژوهش در نظر دارد تا با تعیین آستانه پنج ساله، بیماران سرطان سینه را به عنوان بازماندگان کوتاه مدت یا بازماندگان طولانی مدت طبقه بندی کند. اگر بیماران به عنوان بازماندگان کوتاه مدت طبقه بندی شوند، پزشکان ممکن است بیماران را با درمان‌های تخصصی سرطان توصیه کنند تا از درمان کمکی غیرضروری و رنج ناشی از عوارض جانبی سمی اجتناب کنند. مدل پیش‌بینی این پژوهش سعی خواهد کرد تا از مکانیسم توجه بسته و تکنیک همجوشی اولیه بر روی اطلاعات چندوجهی از جزئیات بالینی و ژنومی (تنوع تعداد نسخه و داده‌های بیان ژن) بیماران مبتلا به سرطان پستان استفاده کند. در حقیقت این پژوهش در نظر دارد تا با بکارگیری الگوریتم شبکه عصبی پیچشی^۱ و تکنیک همجوشی اولیه بر روی اطلاعات چندوجهی به راهکاری مناسب در پیش‌بینی بقای سرطان سینه دست یابد.

مرحله سرطان سینه با حساسیت‌های سرطان تعیین می‌شود، مانند اینکه چقدر بزرگ است و گیرنده‌های هورمونی دارد یا خیر. مرحله سرطان پستان معمولاً به صورت عددی در مقیاس ۰ تا ۵ بیان می‌شود. مرحله ۰ توصیف کننده سرطان‌های غیر تهاجمی است که در محل اصلی خود باقی می‌مانند و مرحله ۵ توصیف کننده سرطان‌های مهاجمی است که در خارج از پستان به سایر قسمت‌های بدن گسترش یافته‌اند. گزارش پاتولوژی شامل اطلاعاتی خواهد بود که برای محاسبه مرحله سرطان پستان استفاده می‌شود. یعنی اینکه آیا این سرطان محدود به یک ناحیه در پستان است یا به بافت‌های سالم داخل سینه یا سایر قسمت‌های بدن گسترش یافته است (۸). سیستم مرحله بندی سرطان سینه که سیستم TNM نام دارد، توسط کمیته مشترک سرطان آمریکا نظارت می‌شود. این کمیته شامل گروهی از متخصصان

¹ Convolutional Neural Network (ConvNet/CNN)

زمان تغییر می کند، ساخته شد. بنابراین، اهمیت متغیرها برای سه دوره زمانی مختلف (یعنی یک، پنج و ده سال) تعیین شد. تحقیق (۱۳) از رویکردهای یادگیری ماشین برای توسعه مدل هایی برای شناسایی و تجسم نشانه های پیش آگهی مربوط به میزان بقای سرطان سینه استفاده کرده است. مقاله (۱۴) یک مدل مجموعه ای انباشته^۷ بهینه سازی شده با فرایارامتر بیزی موازی را پیشنهاد کرده است. این مدل با استفاده از انباشته کردن مدل های یادگیری ماشین، به عنوان مثال، شبکه عصبی عمیق^۸، ماشین تقویت کننده گرادیان^۹ و جنگل تصادفی توزیع شده^{۱۰} توسعه یافته است. مقاله (۱۵) یک رویکرد ترکیبی چندوجهی سلسله مراتبی جدید به نام HFBSurv را با استفاده از مدل دوخطی فاکتوریزه شده برای ترکیب حساسیت های ژنومی و تصویری گام به گام ارائه کرده است. مطالعه (۱۶) با هدف پیش بینی سرطان پستان با استفاده از روش های مختلف یادگیری ماشین با استفاده از داده های دموگرافیک، آزمایشگاهی و ماموگرافی انجام شد. مقاله (۱۷) یک مطالعه مقایسه ای از تکنیک های مختلف یادگیری ماشین برای تشخیص سرطان سینه ارائه می دهد. همچنین یک مدل رمزگذار خودکار را ارائه می دهد که تشخیص سرطان سینه را به شیوه ای بدون نظارت انجام می دهد. در مقاله (۱۸) استفاده از یادگیری ماشین را با استفاده از تصاویر کل اسلاید و اطلاعات بالینی برای پیش بینی پاسخ بافت شناسی به NACT برای بیماران زن اولیه TNBC بررسی نموده است.

روش پیشنهادی:

هدف اصلی این پژوهش ساختن الگویی برای پیش بینی بقای سرطان سینه متمرکز بر الگوریتم شبکه عصبی پیچشی جهت افزایش دقت در پیش بینی با استفاده از مجموعه

روش های درمان را بر اساس مراحل سرطان سینه را شامل مواردی مانند جراحی برای برداشتن توده، هورمون درمانی برای اطمینان از عدم ایجاد سرطان درمان، پرتودرمانی برای درمان و سایر روش ها در نظر گرفت. از اینرو، پیش بینی این بیماری دارای اهمیت فراوانی است.

سرطان سینه دومین عامل مرگ و میر در جهان است. تحقیقات سرطان پستان بر پیش بینی، تشخیص و پیش آگهی اولیه آن متمرکز است. با وجود تمام پیشرفت های انجام شده در پیشگیری و مداخله زود هنگام، پیش آگهی اولیه و نرخ پیش بینی بقا هنوز کافی نیست (۸). یادگیری ماشین یکی از زیر مجموعه های هوش مصنوعی است که به سیستم ها این امکان را می دهد تا به صورت خودکار یادگیری و پیشرفت داشته باشند. الگوریتم های بسیار مختلفی برای یادگیری ماشین وجود دارد و هر روزه صدها الگوریتم جدید نیز تولید می شوند، و به طور معمول توسط سبک یادگیری^۱ مانند یادگیری نظارت شده^۲، یادگیری بدون نظارت^۳، یادگیری نیمه نظارت^۴ و یا با توجه به شباهتشان در فرم و عملکرد مانند طبقه بندی، برگشت، درخت تصمیم گیری، دسته بندی و یادگیری عمیق گروه بندی می شوند. مقاله (۱۰) یک مدل شبکه عصبی مصنوعی^۵ را پیشنهاد کرده که با رسیدن به ۹۹.۵۷٪ از نظر دقت در مجموعه داده سرطان پستان ویسکانسین، از تمام روش های یادگیری نظارت شده قبلی بهتر عمل کرده است. نتایج بهترین نتایج گزارش شده از شبکه عصبی مصنوعی با استفاده از محیط JNN بدون هیچ گونه پیش پردازشی از مجموعه داده است. در مطالعه (۱۱) مدل مبتنی بر ماشین یادگیری افراطی^۶ برای پیش بینی بقای سرطان پستان با نام eBreCaP پیشنهاد شده است. در پژوهش (۱۲) یک روش مبتنی بر داده کاوی ترکیبی برای متمایز کردن متغیرهایی که اهمیت آنها برای بقا در طول

⁷ Bayesian hyperparameter optimized Stacked ensemble (BSense)

⁸ Deep Neural Network (DNN)

⁹ Gradient Boosting Machine (GBM)

¹⁰ Distributed Random Forest (DRF)

¹ Learning Style

² Supervised

³ Unsupervised

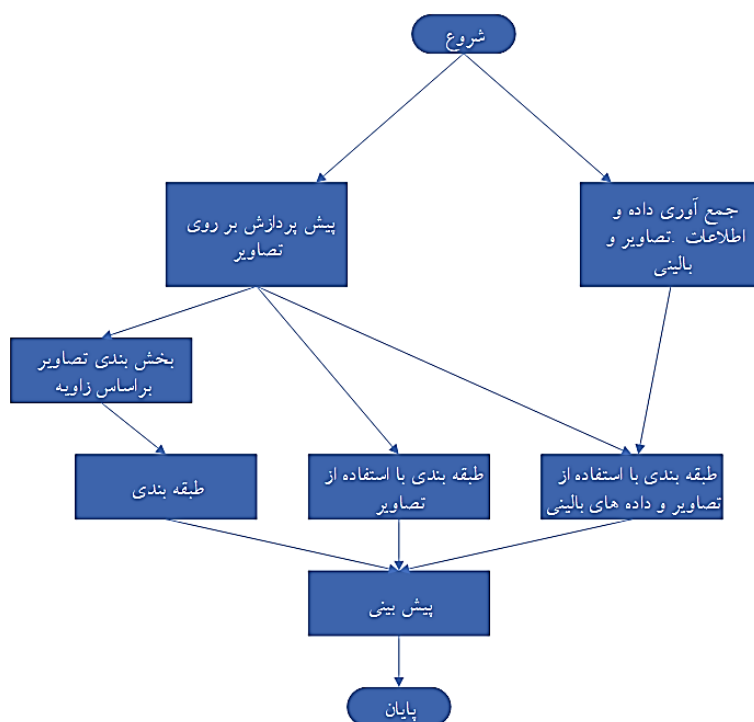
⁴ Semi-Supervised

⁵ Artificial Neural Network (ANN)

⁶ Extreme Learning Machine (ELM)

اگرچه هدف این پژوهش استفاده از تصاویر ماموگرافی است، اما به دلیل اهمیت داده‌های بالینی سعی گردید تا با ترکیب همه داده‌های موجود در مجموعه داده فرایند با دقت و کارایی بیشتری انجام شود. در حقیقت در روش پیشنهادی ابتدا داده‌های مربوط به تصاویر ماموگرافی و اطلاعات بالینی به عنوان ورودی به الگوریتم پیشنهادی وارد خواهد شد.

داده‌های مناسب است. به عبارت دیگر، هدف اصلی این پژوهش ایجاد تصویری واضح از پیش بینی بقای سرطان سینه است. شکل ۱ مراحل اجرای روش پیشنهادی را برای پیش‌بینی بقای سرطان سینه نشان می‌دهد. همانطور که در شکل ۱ مشخص است، مجموعه داده دارای دو بخش تصاویر و داده‌های بالینی مربوط به بیماران است که از هر دو اطلاعات و داده‌ها برای پیش‌بینی استفاده خواهد شد.



شکل ۱- فلوچارت اجرای روش پیشنهادی

که در این حالت یک مرحله اضافی برای آنها اجرا می‌شود و سپس در مرحله بعد از بخش بندی تصاویر، فرآیند طبقه بندی بر روی آنها اجرا می‌شود. حالت دوم مربوط به زمانی است که فرایند طبقه بندی بدون بخش بندی تصاویر انجام می‌شود. حالت سوم، ادغام بین داده‌های بالینی و تصاویر ماموگرافی را انجام می‌دهد که در آن فرآیند طبقه بندی براساس دو بعد داده مختلف انجام می‌شود و اجرای این مرحله اگرچه نسبت به حالت اول یک مرحله کمتر است، یا مراحل آن با حالت دوم یکسان است، اما به دلیل تفاوت در

از آنجایی که داده‌های ورودی در دو بعد یا بخش مختلف هستند، لذا نیاز است تا فرایند اجرای الگوریتم پیشنهادی در ابتدا به صورت جداگانه روی داده‌ها بررسی شود. از این رو، در این مرحله ابتدا برای تصاویر بدست آمده از ماموگرافی، فرآیند یکسان سازی تصاویر انجام شد تا همه آنها به یک اندازه مشخص تبدیل شوند. پس از یکسان سازی تصاویر فرآیند پیش پردازش روی آنها انجام شد بعد از مرحله پیش پردازش، سه حالت برای تصاویر در نظر گرفته شد. حالت اول مربوط به زمانی است که تصاویر باید بخش بندی شوند

داده‌های غیرعادی و پرت یا نویزدار وجود داشت، از مجموعه داده حذف شدند. در حقیقت انجام فرآیند پیش پردازش باعث خواهد شد تا مجموعه داده از وجود داده‌های بدون اطلاعات دقیق و ناکافی تخلیه شود و به نوعی داده‌ها به یک فرمت یکسان برسند.

معماری و فرآیند سیستم پیشنهادی:

در روش این پژوهش، یک شبکه عصبی کانولوشن برای هر یک از پنج نمای سینه که تصاویر گرفته شده از زوایای مختلف توسط دستگاه ماموگرافی بدست آمده، ساخته شد. وجود پنج نما به معنی این نیست که تصاویر توسط نرم افزار یا برنامه ای به زوایای مختلف درجه بندی شده اند و در حقیقت به این معنی است که به منظور تشخیص دقیق تر وجود علائمی از بقای سرطان سینه در فرد بیمار، سعی شده تا تصاویر از زوایای مختلف از افراد و بیماران گرفته شود. در حقیقت به جهت افزایش دقت و کارایی در پیش بینی سعی شد هر مدل به طور جداگانه آموزش داده شود تا اطمینان حاصل گردد که آنها قبل از ترکیب خروجی خود به طور جداگانه عملکرد خوبی دارند. خروجی هر مدل یک مجموعه خواهد بود که شامل پیش بینی وجود بقای سرطان در بیمار یا عدم وجود بقای سرطان سینه و سالم بودن یک بیمار است. خروجی تولید شده توسط هر مدل استفاده شده در شبکه عصبی پیچشی نشان دهنده وجود بقای سرطان، عدم وجود بقای سرطان سینه و سالم بودن فرد خواهد بود. پس از تکمیل آموزش پنج شبکه عصبی پیچشی، نتایج برای آموزش شبکه عصبی نهایی ترکیب می شوند. نتیجه ترکیبی یک مجموعه ۱۰ عنصری خواهد بود به طوری که هر یک از ۳ نما در احتمال وجود بقای سرطان سینه یا احتمال وجود بقای سرطان سینه کمک می کند. خروجی حاصل از شبکه عصبی پیچشی نهایی، پیش بینی توسط مدل چند ورودی CNN خواهد بود. مرحله بعدی اضافه کردن داده‌های شخصی و بالینی به مدل و مشاهده تغییر در عملکرد مدل پس از الحاق نتیجه مدلی است که بر روی داده‌های بالینی آموزش داده شده است. داده‌های بالینی جمع آوری شده

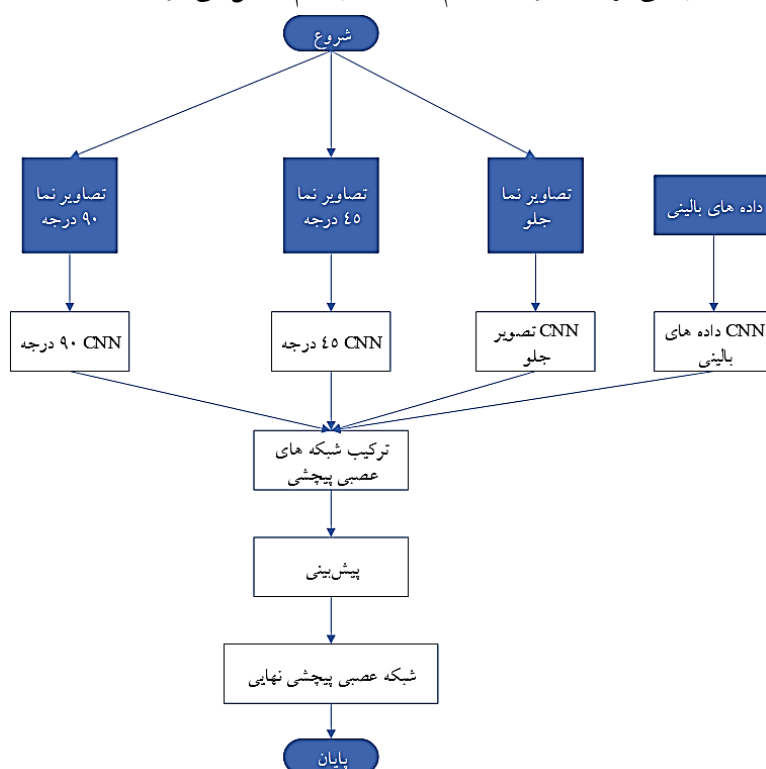
نوع داده‌ها ممکن است مدت اجرای آن کمی طولانی تر از روش‌های دیگر باشد. در نهایت نیز با اتمام حالت‌های ارائه شده، نوبت به اجرای بخش پیش بینی می‌رسد که در فرآیند اجرای الگوریتم پیشنهادی جهت پیش بینی بقای سرطان سینه انجام می‌گیرد.

پیش پردازش مجموعه داده:

پیش پردازش داده گامی در فرآیند داده کاوی و تجزیه و تحلیل داده است که داده‌های خام را می‌گیرد و به قالبی تبدیل می‌کند که توسط رایانه‌ها و یادگیری ماشین قابل درک و تجزیه و تحلیل باشد. داده‌های خام و واقعی در قالب متن، تصویر، ویدیو و غیره نامرتب هستند. نه تنها ممکن است حاوی خطاها و ناهماهنگی باشد، بلکه اغلب ناقص هستند و طراحی منظم و یکنواختی ندارند. ماشین‌ها دوست دارند اطلاعات خوب و مرتب را پردازش کنند، آنها داده‌ها را به صورت صفر و یک تبدیل نمایند. بنابراین محاسبه داده‌های ساخت یافته، مانند اعداد کامل و درصد آسان است. برای افزایش تنوع و تعمیم‌پذیری مدل، از تکنیک‌های داده‌افزایی شامل چرخش، برش و تغییر مقیاس تصاویر ماموگرافی و همچنین افزودن نویز و تنظیم جزئی مقادیر در داده‌های بالینی استفاده شده است. این فرآیندها به صورت جداگانه بر روی هر نوع داده اعمال شده و سپس برای آموزش مدل نهایی ترکیب شده‌اند. این اقدامات به بهبود عملکرد مدل در تشخیص و پیش‌بینی بقای سرطان سینه کمک کرده‌اند. بنابراین با توجه به اهمیت این مرحله، تصاویری که جمع آوری شده بودند مورد بررسی و با توجه به ابعاد متفاوت نیاز بود تا ابتدا اندازه آنها به یک سایز یکسان تبدیل شود. برای این کار، همه تصاویر به ابعاد 640×640 تبدیل شدند. پس از آن تصاویر و داده‌ها مورد بررسی قرار گرفتند و بیمارانی که تصاویر مبهم داشتند یا اینکه تصاویر بدست آمده از آنها از پروتکل پیروی نمی‌کردند و شامل هر پنج نما نبودند، از مجموعه داده حذف شدند. از سوی دیگر، در بخش داده‌های بالینی نیز اطلاعات مربوط به بیماران بررسی شد و موارد که در آنها

دهد. از این رو، در این پژوهش، یک شبکه عصبی پیچشی با پنج لایه (دو لایه پیچشی، دو لایه pooling و یک لایه کاملاً همبند) برای بررسی داده‌های بالینی پیشنهاد گردید. بعد از اینکه این مدل شبکه عصبی بر روی داده‌های بالینی آموزش داده شد، خروجی آن به خروجی مدل چند تصویری CNN اضافه می‌شود که یک مجموعه ۵ عنصری را ارائه می‌دهد که به عنوان ورودی به شبکه عصبی پیچشی نهایی ارسال می‌شود. هم تصاویر و هم داده‌های بالینی به طور جداگانه بررسی می‌شوند و فقط خروجی‌های مدل‌های مربوطه (یعنی CNN برای تصاویر و شبکه عصبی پیچشی برای داده‌های بالینی) برای تعیین طبقه‌بندی نهایی به هم متصل می‌شوند.

شامل اطلاعاتی درباره سن، شرایط، علائم، سابقه خانوادگی، قاعدگی و غیره است. پس از تکمیل طبقه‌بندی این مجموعه داده‌ها، یک شبکه عصبی برای داده‌های بالینی ساخته شد. خروجی این شبکه عصبی مقدراری بین ۰ و ۱ خواهد داشت. آستانه ۰.۵ برای تصمیم‌گیری در مورد پیش‌بینی وجود بقای سرطان سینه و عدم وجود آن تنظیم گردید. از آنجایی که داده‌های بالینی و تصاویر ماموگرافی دو مجموعه داده متفاوت هستند، لذا امکان استفاده از این دو مجموعه در یک شبکه عصبی پیچشی وجود ندارد و باید هر کدام به صورت جداگانه بررسی شوند و شبکه عصبی پیچشی نیز براساس حساسیت‌های متفاوت در تصاویر ماموگرافی و داده‌های بالینی فرآیند خود را انجام



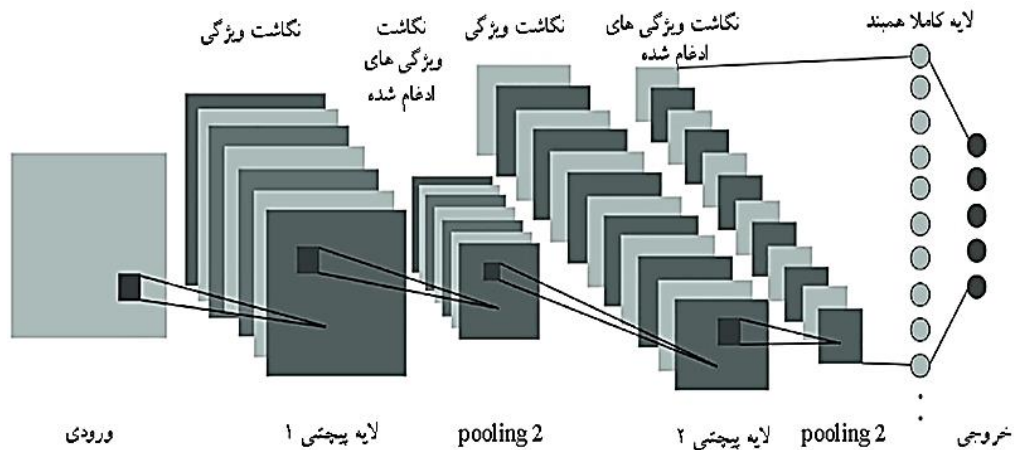
شکل ۲- نحوه پیاده سازی شبکه عصبی پیچشی با چند ورودی

در شبکه عصبی پیچشی نهایی اجرا شدند تا نحوه چگونگی افزودن داده‌های بالینی به مدل و تأثیر آن بر عملکرد مورد بررسی قرار گیرد. شکل ۳، فرآیندی که در شبکه عصبی پیچشی به منظور پیش‌بینی و طبقه‌بندی انجام می‌شود، را

یک نمای کلی از مراحل انجام شده برای به دست آوردن پیش‌بینی با استفاده از تصاویر ماموگرافی به تنهایی و تصاویر ماموگرافی همراه با داده‌های بالینی در شکل ۲ نشان داده شده است. همچنین معیارهای ارزیابی عملکرد متفاوتی

و برای استخراج حساسیت های سطح بالا مانند لبه ها از تصویر ورودی استفاده می شود. لایه بعدی pooling است که مشابه لایه پیچشی، مسئول کاهش اندازه فضایی حساسیت پیچشی شده است.

نشان می دهد. همانطور که در شکل ۳ نشان داده شده است، شبکه عصبی پیچشی دارای پنج لایه مختلف است که هر کدام وظیفه خاصی را برای پیش بینی و طبقه بندی انجام می دهند. لایه اول که لایه پیچشی است، به عنوان لایه هسته



شکل ۳- لایه های شبکه عصبی پیچشی در پیش بینی بقای سرطان سینه

تجزیه و تحلیل نتایج:

مجموعه داده:

پایگاه داده Breast Histopathology Images یک مجموعه داده بزرگ و معتبر است که برای مطالعه و تشخیص سرطان پستان از طریق تحلیل تصاویر بافت شناسی استفاده می شود (۱۹). این مجموعه داده شامل ۱۹۸۷۳۸ قطعه تصویر از بافت های سالم و ۷۸۷۸۶ قطعه تصویر از بافت های مبتلا به سرطان است. تصاویر موجود در این مجموعه، از برش های نازک بافت پستان گرفته شده اند که با رنگ آمیزی هماتوکسیلین و اتوزین آماده سازی شده اند و سپس با استفاده از میکروسکوپ دیجیتال با وضوح بالا تصویربرداری شده اند. این مجموعه داده به دلیل حجم بالای داده ها و تنوع نمونه های موجود، یک منبع ارزشمند برای توسعه و ارزیابی مدل های یادگیری ماشین و یادگیری عمیق در تشخیص و طبقه بندی سرطان پستان محسوب می شود. مدل هایی که بر روی این مجموعه داده آموزش می بینند، قادرند الگوها و ویژگی های مهم در تصاویر بافت شناسی را

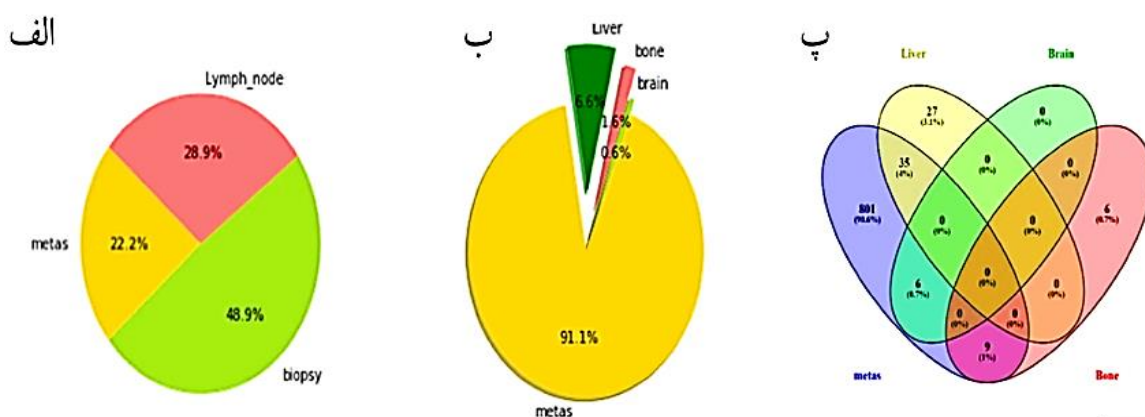
همچنین این لایه وظیفه کاهش توان محاسباتی مورد نیاز برای پردازش داده ها از طریق کاهش ابعاد را دارد. علاوه بر این، برای استخراج حساسیت های غالب نیز مفید است. دو لایه دیگر پیچشی و pooling نیز برای همین فرآیندها مورد استفاده قرار گرفتند. در نهایت برای یادگیری ترکیبات غیرخطی حساسیت های سطح بالا، توسط خروجی لایه پیچشی، یک لایه کاملاً متصل یا همبند در مدل قرار داده شد. بنابراین تصویر ورودی که به فرم مناسب چند سطحی تبدیل شده است، توسط این شبکه به یک بردار ستونی مسطح تبدیل می شود. خروجی مسطح شده به یک شبکه عصبی پیش خور داده می شود و در هر تکرار، الگوریتم پس انتشار به روزرسانی وزن ها را انجام می دهد. س از چندین دوره آموزشی، مدل قادر خواهد بود ویژگی های اصلی و برخی از ویژگی های سطح پایین تر در تصاویر را شناسایی کرده و با استفاده از تابع Softmax آن ها را به دسته های مختلف تخصیص دهد.

پرونده پزشکی و (۲) گزارش هیستوپاتولوژیک با مجموع ۲۵۶۵۲ ورودی داده جاسازی شده است. همانطور که در بخش پیش پردازش اشاره شد.

شناسایی کرده و به تشخیص دقیق تر و سریع تر این نوع سرطان کمک کنند. در کنار این تصاویر، اطلاعات پاتولوژیک جمع آوری شده آن نیز، در دو ستون (۱) شماره

S.NO	NO OF CHEMO CYCLE	HB % gms	Red Cell Count 10(12)/L	WBC Count 10(9)/L	Neutrophils (%)	Lymphocys (%)	Monocyte (%)
1	Base line	12.8	4.75	7.53	62	32	5
	AED_C1	13	4.84	4.66	52	35	11
	AED_C2	13.2	4.89	6.74	64	24	10
	AED_C3	11.4	4.18	5.82	64	20	16
	AED_C4	12	4.26	6.57	64	25	10
	Taxel_C1	12.9	4.57	7.48	66	24	9
	Taxel_C2	12.5	4.5	5.92	67	25	7
	Taxel_C3	12.8	4.68	8.36	75	18	7
	Taxel_C4	13.3	5.03	7.81	77	15	8
2	Base line	13	4.86	8.6	71	22	6
	AED_C1	12	4.47	9.72	79	14	7
	AED_C2	11.7	4.31	8.35	78	14	8
	AED_C3	11.5	4.18	7.68	76	14	10
	AED_C4	11.4	4.1	6.78	72	18	10
	Taxel_C1	11.3	4.04	4.11	63	25	10
	Taxel_C2	12.1	4.22	5.58	72	20	8
	Taxel_C3	11	3.84	4.66	71	18	10
	Taxel_C4	13.6	5.05	9.26	60	35	4
3	Base line	12.6	4.61	8.84	58	32	9
	AED_C1	12.6	4.58	5.24	50	39	10
	AED_C2	11.8	4.29	7.8	63	23	12
	AED_C3	12	4.33	7.33	70	21	9
	AED_C4	12.5	4.5	10.61	75	18	7
	Taxel_C1	12.1	4.32	8.83	66	22	10
	Taxel_C2	12.4	4.48	7.16	54	34	8
	Taxel_C3	12.2	4.63	7.64	52	38	7

شکل ۴- نمونه ای از رکوردهای مجموعه داده.



شکل ۵- (الف) نمودار دایره ای که نسبت بیماری که متاستاز و درگیری غدد لنفاوی دارند. (ب) نمودار دایره ای که درصد سلول های سرطانی منتشر شده در کبد، استخوان و مغز بیماران مبتلا به سرطان سینه متاستاتیک. (پ) نمودار ون نشان دهنده تعداد سوابق پزشکی متقاطع بین متاس و استخوان، مغز و کبد.

متاستاز استفاده می شود. از بین ۵۱۷۶ نمونه موجود در مجموعه داده ۱۸۷۵ مورد با بیوپسی، ۱۱۱۰ با غدد لنفاوی و ۸۵۱ با متاستاز مطابقت داشتند. شکل ۵ (الف) نمودار دایره ای مربوط به درصد سوابق پزشکی که با بیوپسی، غدد لنفاوی و متاستاز مطابقت دارند، را نشان می دهد. از بین ۸۵۱

مجموعه داده بعد از پردازش اولیه، به ۵۱۷۶ ورودی کاهش یافته و به عنوان یک فایل Breast.csv ذخیره گردید که در شکل ۴ برخی از جزئیات این فایل ارائه شده است (۱۹). این فایل برای شناسایی سوابق پزشکی منطبق با دانش اولیه مرتبط با بدخیمی سرطان، مانند بیوپسی، غدد لنفاوی و

استخوان، ۶ پرونده پزشکی متقاطع بین سرطان سینه و مغز، و ۲۷ پرونده منحصراً در کبد یافت می شود (شکل ۵ (پ)).

۴-۴- معیارهای ارزیابی

برای ارزیابی روش پیشنهادی، معیارهای فراخوانی، دقت، صحت، امتیاز F، حساسیت و نرخ خطا مورد استفاده قرار گرفت که روابط آنها در این بخش شرح داده شده است.

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F - Score = \frac{2 * Recall * Precision}{Recall + Precision} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$Error Rate = 100 - Accuracy \quad (6)$$

نمونه، تعداد ۶۲ مورد MR با کبد، ۱۵ مورد MR با استخوان و ۶ مورد MR با مغز مطابقت داشتند که در شکل ۵ (ب) نشان داده شده است. در حالی که ۷۸۹ مورد MR با درگیری غدد لنفاوی مطابقت داشتند. برای شناسایی بیماران مبتلا به بقای سرطان سینه، مجموعه داده با استفاده از نمودار Venn مقایسه شده که در آن ۳۵ پرونده پزشکی وجود دارد که بین سرطان سینه و کبد، ۹ پرونده پزشکی بین سرطان سینه و

مصنوعی (ANN) (۱۶)، الگوریتم بهینه سازی زنبورهای عسل مصنوعی (ABC) (۱۴)، مدل ترکیبی یادگیری ماشین (شامل درخت تصمیم، چنگل تصادفی، بیز ساده، رگرسیون لجستیک و k-نزدیکترین همسایه) (MML) (۲۰)، الگوریتم یادگیری ماشین مبتنی بر درخت تصمیم (DT) (۱۸) است که براساس معیارهای اشاره شده در بخش قبل ارزیابی شده اند. پیاده سازی مدل پیشنهادی و سایر روش های مقایسه ای در این مطالعه با استفاده از یک سیستم مجهز به پردازنده Intel Core i7، ۱۶ گیگابایت حافظه RAM و یک هارد ۵۰۰ گیگابایتی از نوع SSD انجام شده است. سیستم عامل این محیط Windows 10 نسخه ۶۴ بیتی بود و تمام کدهای مربوط به پیاده سازی مدل ها با استفاده از Python نسخه ۳.۹.۱ نوشته و اجرا شده اند. این محیط محاسباتی به گونه ای انتخاب شد که توانایی پردازش داده ها به طور کارآمد را فراهم کند و در عین حال قابلیت اجرای مدل های یادگیری عمیق را داشته باشد. همچنین هاپر پارامترهای شبکه CNN، شامل تعداد فیلتراسیون ها (۶۴، ۱۲۸، ۲۵۶)، اندازه کرنل ها

که در آن TP به عنوان تعداد پیش بینی های که به طور دقیق به عنوان عدم وجود بقای سرطان سینه طبقه بندی شده، TN تعداد پیش بینی که به عنوان وجود بقای سرطان سینه طبقه بندی شده، FP تعداد موارد وجود بقای سرطان سینه که به اشتباه به عنوان عدم وجود بقای سرطان سینه طبقه بندی شده و FN تعداد موارد عدم وجود بقای سرطان سینه که به اشتباه به عنوان وجود بقای سرطان سینه طبقه بندی شده اند، می باشند.

تجزیه و تحلیل نتایج

توانایی یک مدل برای پیش بینی بقای سرطان سینه براساس پارامترهای خاصی انجام می شود تا آن مدل با دیگر مدل ها قابل مقایسه نیز باشد. تمام یافته های عملکرد ارائه شده در این پژوهش، مقدار متوسط خروجی های ۱۰۰ دوره از پیاده سازی روش ها است که برای دو حالت تعداد حساسیت ثابت و متغیر برای مجموعه داده ها اجرا شده است. این پژوهش عملکرد رویکرد پیشنهادی را با چندین تکنیک پیشرفته دیگر که در مطالعات اخیر ارائه شده اند را مقایسه نموده است. این مطالعات و الگوریتم ها شامل مدل شبکه عصبی

(۳×۳)، نرخ یادگیری (۰.۰۰۱)، و اندازه batch (۳۲) می‌باشد که از طریق فرآیندهای cross-validation تنظیم شده‌اند. در این مطالعه، نمونه‌ها به سه مجموعه مجزا برای آموزش، تست و اعتبارسنجی تقسیم‌بندی شدند. به منظور اطمینان از تعمیم‌پذیری مدل و جلوگیری از بیش‌برازش، 70% از داده‌ها به عنوان مجموعه آموزش، 15% به عنوان مجموعه تست و 15% باقی‌مانده به عنوان مجموعه اعتبارسنجی در نظر گرفته شد. مجموعه آموزش برای بهینه‌سازی پارامترهای مدل و یادگیری الگوهای موجود در داده‌ها استفاده شده است. پس از آن، مدل بر روی مجموعه تست ارزیابی شده تا عملکرد آن بر روی داده‌های جدید و دیده‌نشده سنجیده شود. نهایتاً، مجموعه اعتبارسنجی برای ارزیابی نهایی مدل و انتخاب بهترین پارامترها استفاده گردیده است. همچنین، برای اعتبارسنجی نهایی، از روش اعتبارسنجی متقابل^۱

استفاده شده است تا پایداری و قابلیت اطمینان مدل در مواجهه با داده‌های مختلف ارزیابی گردد. این روش به ما کمک کرده است تا از نتایج به‌دست‌آمده اطمینان حاصل کنیم و مدل پیشنهادی را به طور دقیق‌تر با دیگر مدل‌ها مقایسه کنیم. نتایج مدل‌های وجود و عدم وجود بقای سرطانی در سینه در مجموعه داده با در نظر گرفتن داده‌های بالینی و بدون در نظر گرفتن آن در جدول ۱ و ۲ ارائه شده است. جدول ۱ نتایج طبقه‌بندی مربوط به مجموعه داده با در نظر گرفتن داده‌های بالینی را نشان می‌دهد. نتایج نشان می‌دهد که مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی با رویکرد انتخاب حساسیت بهترین نتایج را نسبت به دیگر مدل‌ها به دست آورده است.

جدول ۱- نتایج مدل‌های مختلف برای مجموعه داده با داده‌های بالینی

مدل	حساسیت	صحت	فراخوانی	امتیاز F1	دقت	نرخ خطا
مدل پیشنهادی	۹۶/۳۴	۹۸/۵۶	۹۶/۵۴	۹۸/۵۴	۹۸/۳۴	۱/۶۶
ANN	۹۳/۵۲	۹۲/۳۲	۹۰/۳۲	۹۱/۲۳	۹۱/۹۹	۸/۰۱
ABC	۸۴/۸۷	۸۶/۳۴	۸۴/۳۲	۸۲/۲۳	۸۳/۷۷	۱۶/۲۳
MML	۸۶/۸۱	۸۴/۳۲	۸۲/۱۲	۸۴/۷۸	۸۲/۳۳	۱۷/۶۷
DT	۸۲/۱۲	۷۸/۶۵	۷۸/۰۹	۷۷/۹۸	۷۶/۵۴	۲۳/۴۶

مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی مقادیر ۹۶/۳۴، ۹۸/۵۶، ۹۶/۵۴، ۹۸/۵۴ و ۹۸/۳۴ را به ترتیب برای معیارهای حساسیت، صحت، فراخوانی، امتیاز F و دقت به دست می‌آورد. این نتایج در مقایسه با مدل‌های موجود مقادیر بالاتری را نشان می‌دهد. جدول ۲ نتایج مدل پیشنهادی و دیگر الگوریتم‌ها را برای مجموعه داده بدون در نظر گرفتن داده‌های بالینی مربوط به افراد و بیماران را نشان می‌دهد. مدل پیشنهادی توانسته بالاترین

دقت را نسبت به الگوریتم‌های مورد مقایسه بدست آورد. به همین ترتیب، با مقایسه نتایج نرخ خطا می‌توان دریافت که مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی به نرخ خطای کمتری نسبت به تکنیک‌های موجود رسیده است. نتایج بدست آمده در این دو جدول، تشخیص دقیق‌تر وجود و عدم وجود بقای سرطان سینه را با استفاده از مدل پیشنهادی با دقت بالاتر و نرخ خطای کمتر تأیید می‌کند. برای نمونه‌های عدم وجود بقای

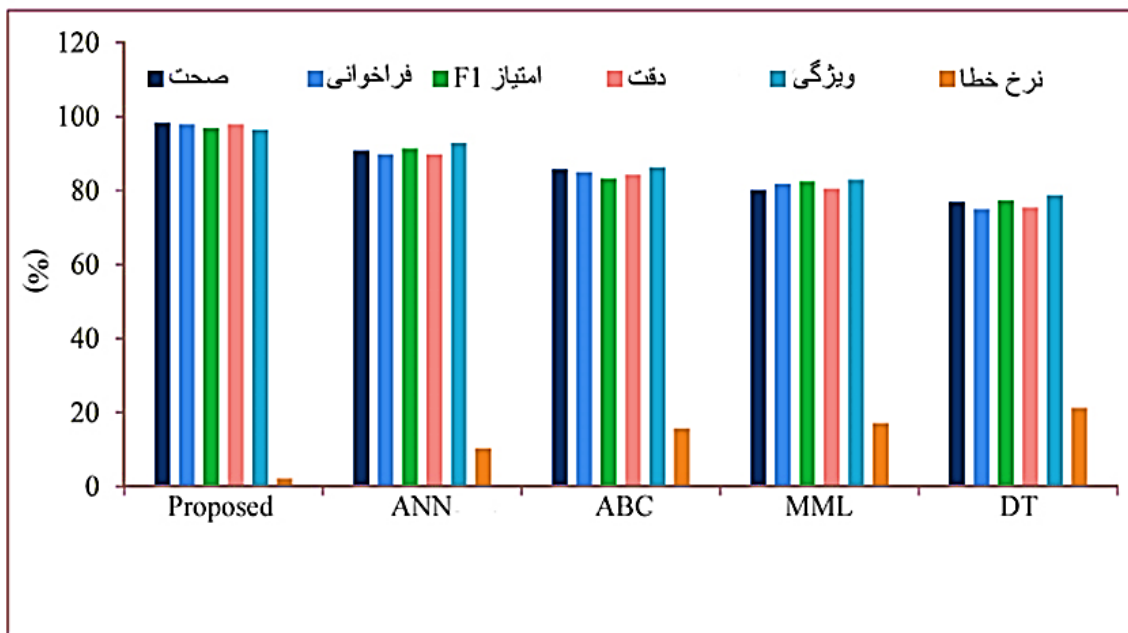
¹ Cross-validation

مدل پیشنهادی است. به همین ترتیب، مدل پیشنهادی با رویکرد انتخاب حساسیت مبتنی بر شبکه عصبی پیچشی با سه لایه کانولوشن بهترین عملکرد را در طبقه بندی و پیش بینی بقای سرطان سینه برای معیارهای مورد هدف پژوهش به دست آورده است.

سرطان سینه در مجموعه داده، مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی به دقت ۹۷/۲۱ درصد دست یافته است. در مقابل الگوریتم های مورد مقایسه شامل ANN، ABC، MML و DT به ترتیب به دقت ۹۰/۲۱، ۸۵/۲۱، ۸۳/۴۵ و ۷۹/۸۷ درصد را به دست آورده اند که کمتر از

جدول ۲- نتایج مدل های مختلف برای مجموعه داده با داده های بالینی

مدل	حساسیت	صحت	فراخوانی	F1 امتیاز	دقت	نرخ خطا
مدل پیشنهادی	۹۷/۰۲	۹۸/۲۲	۹۶/۲۶	۹۶/۱۲	۹۷/۲۱	۲/۷۹
ANN	۹۲/۱۱	۹۰/۲۱	۸۸/۰۹	۸۹/۹۸	۹۰/۲۷	۹/۷۳
ABC	۸۸/۷۶	۸۵/۲۱	۸۴/۲۳	۸۵/۷۶	۸۵/۰۹	۱۴/۹۱
MML	۸۵/۹۲	۸۳/۴۵	۸۱/۹۰	۸۴/۲۲	۸۳/۳۰	۱۶/۷
DT	۸۱/۱۳	۷۹/۸۷	۷۷/۹۸	۷۸/۳۲	۷۸/۹۸	۲۱/۰۲



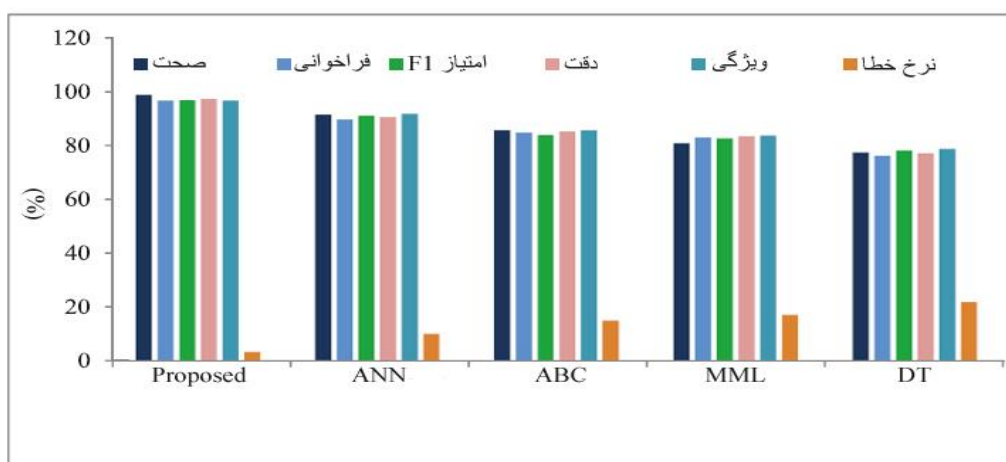
شکل ۶- نتایج مدل های مختلف برای مجموعه داده با تعداد حساسیت متغیر بدون داده های بالینی

تحلیل قرار گرفت. شکل های ۶ و ۷ نتایج حاصل از پیاده سازی مدل پیشنهادی را به همراه الگوریتم های دیگر برای معیارهای حساسیت، صحت، فراخوانی، امتیاز F1، دقت و نرخ خطا را برای موارد وجود و عدم وجود بقای سرطان سینه را نشان داده اند. مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی با انتخاب حساسیت متغیر بدون

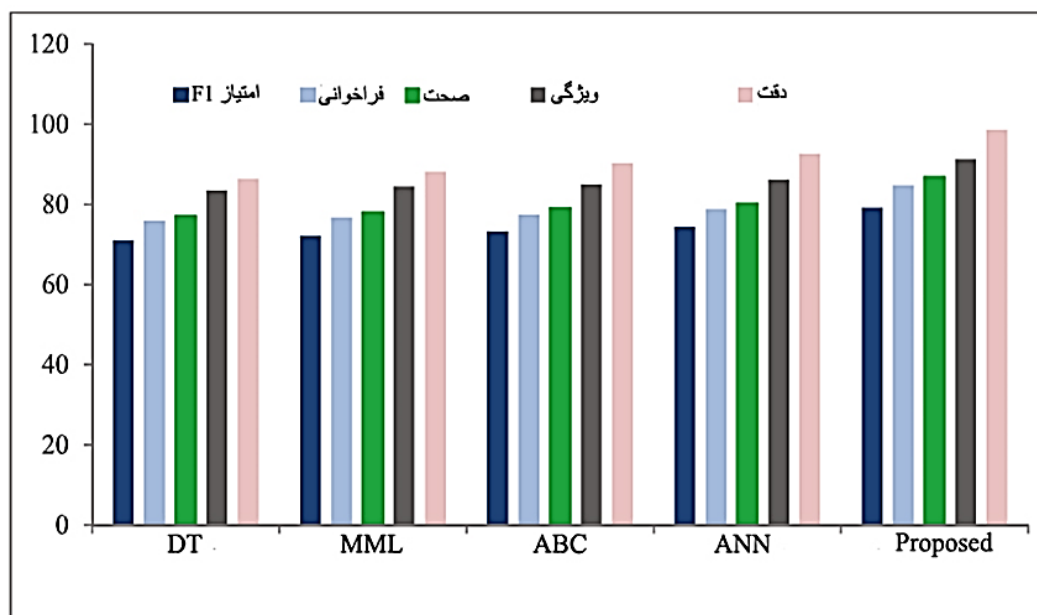
در بخش دیگری از پژوهش، نتایج عملکرد مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی برای مجموعه داده با تعداد حساسیت های متغیر (۲۰ و ۲۴ حساسیت) و با در نظر گرفتن داده های بالینی و بدون آنها با کمک معیارهای عملکرد ارائه شده در بخش قبل با الگوریتم های دیگر در شرایط یکسان مورد تجزیه و

یادگیری ماشین (شامل درخت تصمیم، چنگل تصادفی، بیز ساده، رگرسیون لجستیک و k-نزدیکترین همسایه) و الگوریتم یادگیری ماشین مبتنی بر درخت تصمیم به ترتیب به دقت ۸۹/۷۶، ۸۴/۳۲، ۸۰/۴۳ و ۷۵/۴۵ درصد دست یافته اند که کمتر از مدل پیشنهادی است.

داده‌های بالینی و با در نظر گرفتن داده‌های بالینی به دقت ۹۷/۸۸ درصد برای موارد وجود بقای سرطان سینه دست یافته است که نسبتاً بالاتر از الگوریتم‌های مورد مقایسه است. نتایج بدست آمده از مدل‌های مورد مقایسه، شامل مدل شبکه عصبی مصنوعی، الگوریتم بهینه‌سازی زنبورهای عسل مصنوعی، مدل ترکیبی



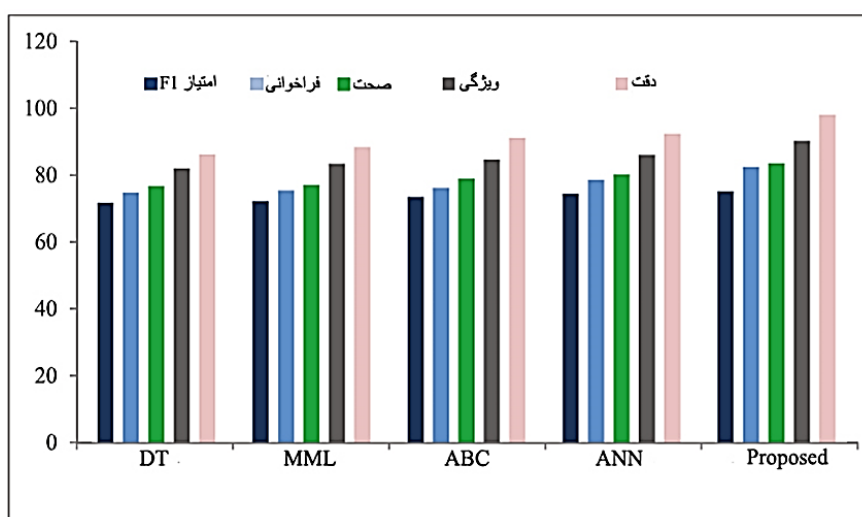
شکل ۷- نتایج مدل‌های مختلف برای مجموعه داده با تعداد حساسیت متغیر با داده‌های بالینی



شکل ۸- نتایج مدل‌های مختلف برای مجموعه داده با تعداد حساسیت ثابت بدون داده‌های بالینی

حساسیت ثابت بدون در نظر گرفتن داده ای بالینی حدود ۶ درصد، دقت بالاتری دارد. همانطور که در شکل های ۸ و ۹ نشان داده شد. مدل پیشنهادی مبتنی بر شبکه عصبی پیچشی در مقایسه با مدل های موجود دقت بیشتری را برای پیش بینی به دست آورده است. از این رو، می توان گفت که مدل پیشنهادی بهترین عملکرد را برای همه طرح های انتخاب حساسیت موارد وجود و عدم وجود بقای سرطان سینه دارد، زیرا مدل پیشنهادی از شبکه عصبی پیچشی با لایه های کانولوشن، ادغام و کاملاً متصل مشخص در فرآیند طبقه بندی استفاده می کند که مشکل گرادیان ناپدید شدن و خطاهای همگرایی را حذف می کند و دقت طبقه بندی را بهبود می بخشد. این نتیجه برای مجموعه داده با تعداد حساسیت ثابت و با در نظر گرفتن داده های بالینی نیز صدق می کند و همانطور که در شکل ۹ نشان داده، مدل پیشنهادی توانسته با دقت و کارایی بالاتری پیش بینی را انجام دهد.

به طور مشابه، مدل های موجود ضعیف ترین مقادیر برای معیارهای حساسیت، صحت، فراخوانی، امتیاز F_1 ، دقت و نرخ خطا را برای موارد وجود بقای سرطان سینه در مجموعه داده بدست آورده اند. شکل های ۸ و ۹ میانگین دقت بدست آمده توسط مدل طبقه بندی کننده پیشنهادی و الگوریتم های مورد مقایسه برای رویکردهای یادگیری حساسیت های ثابت براساس دو حالت در نظر گرفتن داده های بالینی و بدون داده های بالینی را نشان می دهد. دقت مدل های مختلف برای مجموعه داده با حساسیت ثابت بدون در نظر گرفتن داده های بالینی در شکل ۸ نشان داده شده است. با رویکرد انتخاب حساسیت ثابت و بدون در نظر گرفتن داده های بالینی، مدل پیشنهادی به دقت ۹۸/۵۶ درصد دست یافته است که نسبت به مدل های دیگر نتایج بهتری را به همراه دارد و برای DT، ABC و ANN دقت به ترتیب برابر با ۸۸/۴۱، ۹۲/۵۶ و ۹۲/۵۶ درصد است. این نتایج نشان می دهد که مدل پیشنهادی نسبت به حالت انتخاب



شکل ۹- نتایج مدل های مختلف برای مجموعه داده با تعداد حساسیت ثابت با داده های بالینی.

نتیجه گیری

از برخی حساسیت ها جهت پیش بینی بقای سرطان سینه بکار گرفته شده است که در آن با افزودن برخی مراحل به فرآیند پیش پردازش و آموزش سعی شده است تا میزان دقت

از آنجایی که هدف اصلی الگوریتم پیشنهادی پیش بینی بقای سرطان سینه است. برای این منظور در این پژوهش از مدل یادگیری عمیق شبکه عصبی پیچشی به همراه استفاده

راحتی می‌تواند در کاربردهای مختلف جهت پیش بینی بیماری‌ها مانند سرطان سینه، سرطان ریه و غیره استفاده شود. الگوریتم پیشنهادی سعی در استفاده از روشی داشته است که بتواند دقت را در پیش بینی افزایش دهد و از سوی دیگر نتایج پیش بینی را به اطلاعات واقعی نزدیک نماید به نحوی که اختلاف به کمترین مقدار برسد.

الگوریتم پیشنهادی افزایش یافته و کارایی آن جهت انجام فرآیند پیش بینی افزایش یابد. در این پژوهش از مدل شبکه عصبی پیچشی به منظور استخراج حساسیت‌های موثر و با اهمیت در پیش بینی بقای سرطان سینه استفاده شده و سپس الگوریتم زوایا و نماهای مختلف تصاویر برای پیش بینی دقیق و تقویت بیشتر دقت الگوریتم استفاده شده است که به

References

1. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep*. 2021;11(1):6968.
2. Li J, Zhou Z, Dong J, et al. Predicting breast cancer 5-year survival using machine learning: A systematic review. *PLoS One*. 2021;16(4):e0250370.
3. Okagbue HI, Adamu PI, Oguntunde PE, et al. Machine learning prediction of breast cancer survival using age, sex, length of stay, mode of diagnosis and location of cancer. *Health Technol*. 2021; 11, 887–893.
4. Gupta S, Manoj KG. A comparative analysis of deep learning approaches for predicting breast cancer survivability. *Archives of Computational Methods in Engineering*. 2022; 29(5): 2959-2975.
5. Xu Y, Peng Y, Shen M, et al. Construction and Validation of Angiogenesis-Related Prognostic Risk Signature to Facilitate Survival Prediction and Biomarker Excavation of Breast Cancer Patients. *J Oncol*. 2022;2022:1525245.
6. Ravale U, Yashashree B. Breast Cancer Prediction Using Different Machine Learning Algorithm. In *Sentiment Analysis and Deep Learning*. 2023; 493-502.
7. Zhang J, Guihong L, Zili D, Fuchuan X, Ronghui Z, Benchao Y, Liyi G. Novel RNA N6-methyladenosine regulator related signature for predicting clinical and immunological characteristics in breast cancer. 2023; (2023): 147095.
8. Kai z, Di S, Li S, Liang X, Zhai Y, Chen J, Ouyang K, Cappello F, Chen Z. FT-CNN: Algorithm-based fault tolerance for convolutional neural networks. *IEEE Transactions on Parallel and Distributed Systems*. 2020; 32(7): 1677-1689.
9. VirenViraj S, Kumar V, Devagade U, Karanth V. Heart disease prediction using CNN algorithm. *SN Computer Science*. 2020; 1 (3): 1-8.
10. Shawarib, Mohammed Ziyad Abu, Ahmed Essam Abdel Latif, Bashir Essam El-Din Al-Zatmah, and Samy S. Abu-Naser. Breast Cancer Diagnosis and Survival Prediction Using JNN. *International Journal of Engineering and Information Systems (IJEAIS)*. 2020; 4(10).
11. Dhillon, Arwinder, and Ashima Singh. eBreCaP: extreme learning-based model for breast cancer survival prediction. *IET Systems Biology*. 2020; 14(3): 160-169.
12. Serhat S, Kursuncu U, Kibis E, AnisAbdellatif M. A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival. *Expert Systems with Applications*. 2020; 139: 112863.
13. Mohammad Nazmul H, Tazin T, Monirujjaman Khan M, Faisal S, Md Ibraheem S, Algethami H. Predicting Characteristics Associated with Breast Cancer Survival Using Multiple Machine Learning Approaches. *Computational and Mathematical Methods in Medicine*. 2022.
14. Parampreet K, Singh A, Chana I. BSense: A parallel Bayesian hyperparameter optimized Stacked ensemble model for breast cancer survival prediction. *Journal of Computational Science* 60. 2020; 101570.
15. Ruiqing L, Wu X, Ao Li, Wang M. HFBSurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction. *Bioinformatics*. 2022; 38(9):2587-2594.
16. Reza R, Ayyoubzadeh SM, Sohrabei S, Esmaili M, Atashi A. Prediction of Breast Cancer using Machine Learning Approaches. *Journal of Biomedical Physics and Engineering*. 2022; 12(3): 297-308.
17. Sonam Jawahar S, Rajaraman R, Verlekar T. Breast Cancer Prediction Using Auto-Encoders. In *International Conference on Data Management, Analytics & Innovation*. 2023; 121-132.

18. Ogier du Terrail J, Leopold A, Joly C, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat Med.* 2023;29(1):135-146.
19. Botlagunta M, Botlagunta MD, Myneni MB, et al. Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. *Sci Rep.* 2023;13(1):485.
20. Singh D, Nigam R, Mittal R, et al. Information retrieval using machine learning from breast cancer diagnosis. *Multimed Tools Appl.* 2023; 82, 8581–8602.

Original Article

Prediction of breast cancer survival using deep learning network and multivariate data

Received: 28/03/2024 - Accepted: 09/09/2024

Fatemeh Jalali¹

Mohammad Mehdi Hosseini*²

1 Department of Engineering, Shahrood Institute of Higher Education (non-profit), Shahrood, Iran

2 Department of Engineering, Shahrood Branch, Islamic Azad University, Shahrood, Iran

Email:

Hosseini_mm@shahroodut.ac.ir

Abstract

Introduction

Breast cancer, as one of the most common and significant cancers, plays a major role in increasing mortality rates among women. Currently, multi-modal cancer-related data, including genomic details, mammography images, and clinical information, are available, which has led to increased focus on developing advanced deep-learning models for predicting breast cancer survival. This study proposes a deep learning model based on convolutional neural networks (CNN) for predicting breast cancer survival using multi-modal data. This model extracts crucial features to enhance prediction accuracy. The proposed method comprises two phases: in the first phase, a CNN is used to extract effective features, and in the second phase, these features are employed to predict breast cancer survival. The experimental results demonstrated that the proposed method achieved an accuracy of 98.34% when clinical data were utilized, and 97.21% without clinical data. This indicates that the proposed model significantly outperforms the compared models. Moreover, the results show that incorporating clinical data is instrumental in improving the accuracy and performance of the proposed model.

Key words

Breast cancer, mammography, breast cancer survival prediction, deep learning, convolutional neural network (CNN).

Acknowledgement: There is no conflict of interest