

تشخیص وضعیت بیماران مبتلا به هپاتیت C با استفاده از الگوریتم یادگیری

عمیق CNN

کاظم تقندیکی

استادیار گروه مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه فنی و حرفه ای، تهران، ایران

ktaghandiki@tvu.ac.ir

تاریخ پذیرش: ۱۴۰۳/۰۵/۱۳

تاریخ دریافت: ۱۴۰۳/۰۲/۳۱

چکیده

یادگیری عمیق، به عنوان یکی از زیرشاخه‌های مهم هوش مصنوعی، قابلیت‌های فراوانی برای ارتقای سلامت انسان دارد. در سال‌های اخیر، این فناوری به طور گسترده‌ای برای تحلیل، پیش‌بینی و بهبود درمان‌های بیماران مورد توجه قرار گرفته است. در این مطالعه، از الگوریتم یادگیری عمیق شبکه عصبی پیچشی (CNN) و روش‌های پیش‌پردازش داده‌ها با استفاده از زبان برنامه‌نویسی پایتون، برای پیش‌بینی وضعیت مرگ یا زنده‌گی بیماران مبتلا به هپاتیت C استفاده شده است. هدف اصلی این پژوهش، کمک به پزشکان در اتخاذ تصمیمات درمانی دقیق‌تر و مؤثرتر است. نتایج به دست آمده از این پژوهش نشان می‌دهد که رویکرد پیشنهادی با دقت ۹۸٪ قادر به پیش‌بینی وضعیت بیماران مبتلا به هپاتیت C است. این دقت بالا، توانایی الگوریتم‌های یادگیری عمیق در تحلیل داده‌های حجیم و ارائه نتایج قابل اعتماد را به خوبی نشان می‌دهد. بر این اساس، استفاده از الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی پیچشی می‌تواند نقش مهمی در پیش‌بینی و شخصی‌سازی درمان بیماران داشته باشد. این امر می‌تواند به طور قابل توجهی به بهبود بهداشت و سلامت جامعه کمک کند، چرا که با دقت بالاتری می‌توان وضعیت بیماران را پیش‌بینی و درمان‌های مناسب‌تری ارائه داد.

کلیدواژه‌ها: یادگیری ماشین، یادگیری عمیق، هوش مصنوعی، الگوریتم CNN، هپاتیت C

مقدمه

یادگیری عمیق با استفاده از شبکه‌های عصبی مصنوعی که از مغز انسان الهام گرفته شده‌اند، قادر به یادگیری از داده‌های حجیم و پیچیده و انجام وظایف مختلف با دقت بالا است (لی و لو، ۲۰۲۰). این فناوری در حال حاضر در حوزه‌های مختلفی از جمله بینایی رایانه، پردازش زبان طبیعی، رباتیک و تشخیص پزشکی به طور گسترده استفاده می‌شود (گو و آلمتن، ۲۰۲۰). در سال‌های اخیر، یادگیری عمیق به عنوان ابزاری قدرتمند برای کشف و توسعه داروها و درمان‌های جدید نیز ظهور کرده است. یادگیری عمیق می‌تواند در هر یک از مراحل زیر بهبود هایی ایجاد کند و فرآیند کشف و توسعه دارو و همچنین تشخیص بیماری‌ها را تسریع نماید.

۱. شناسایی هدف: یادگیری عمیق می‌تواند با تحلیل داده‌های بیولوژیکی و ژنتیکی، مولکول‌ها و پروتئین‌های خاصی را که به عنوان هدف‌های دارویی بالقوه برای درمان هپاتیت C شناخته می‌شوند، شناسایی کند. این روش به دقت بیشتری نسبت به روش‌های سنتی منجر می‌شود (لیو و لی، ۲۰۲۰).

۲. غربالگری مولکول‌ها: شبکه‌های عصبی عمیق می‌توانند با تحلیل داده‌های شیمیایی، مولکول‌هایی را که توانایی تعامل با اهداف دارویی هپاتیت C را دارند، با سرعت و دقت بالا غربال کنند. این امر باعث می‌شود که تعداد زیادی مولکول به سرعت ارزیابی و مولکول‌های کارآمدتر انتخاب شوند (مسترز و گرگوری-پویجانه، ۲۰۲۰؛ براون و سامانتا، ۲۰۱۹).

۳. آزمایش‌های پیش‌بالینی: یادگیری عمیق می‌تواند با تحلیل داده‌های آزمایش‌های سلولی و حیوانی، ایمنی و اثربخشی مولکول‌های منتخب را پیش‌بینی کند و به کاهش زمان و هزینه‌های این مرحله کمک کند (چن و وونگ، ۲۰۱۹؛ چینگ و همکاران،

.(۲۰۲۰)

۴. آزمایش‌های بالینی: در مراحل بالینی، یادگیری عمیق می‌تواند با تحلیل داده‌های بیمار و پیش‌بینی پاسخ‌های فردی به درمان، به بهینه‌سازی دوزها و روش‌های درمانی کمک کند، که این امر به افزایش اثربخشی و کاهش عوارض جانبی منجر می‌شود (مسترز و گرگوری-پویجانه، ۲۰۲۰).

با ترکیب این مراحل، یادگیری عمیق می‌تواند به طور موثری در کشف و توسعه داروهای جدید برای هپاتیت C و بهبود درمان‌های موجود کمک کند، که نهایتاً منجر به ارتقای سلامت و بهداشت جامعه می‌شود.

بیماری‌های متعددی بر کبد انسان تأثیر می‌گذارند، یکی از مهم‌ترین آن‌ها هپاتیت C است که توسط ویروس هپاتیت (HCV) ایجاد می‌شود و در صورت عدم شناسایی، می‌تواند منجر به مرگ شود (وانگ و همکاران، ۲۰۲۰). در برخی موارد، این ویروس ممکن است ۱۰ الی ۲۰ سال در بدن باقی بماند بدون اینکه فعال شود (ساجدوا و همکاران، ۲۰۲۳). تنها ۳۰٪ از بیماران به‌طور خودبخود در عرض شش ماه بهبود می‌یابند، در حالی که بیشتر آن‌ها به عفونت مزمن مبتلا می‌شوند. دو مشکل عمده ناشی از عفونت مزمن هپاتیت C، سیروز و سرطان کبد است (علی‌زارگر و همکاران، ۲۰۲۳؛ رسلی و همکاران، ۲۰۱۸؛ ممدوح و همکاران، ۲۰۲۲).

اغلب مبتلایان به هپاتیت C علائمی نشان نمی‌دهند و علائم زمانی ظاهر می‌شوند که بیماری پیشرفت کرده باشد. بیماران ممکن است احساس ضعف، خواب‌آلودگی و سرگیجه کنند که با خستگی ناشی از کار یا مطالعه اشتباه گرفته می‌شود. این موضوع نگران‌کننده‌تر می‌شود وقتی بدنیم واکسن هپاتیت C هنوز در دسترس نیست (استوکس و یانگ، ۲۰۲۰). هپاتیت C یک بیماری شایع است که بیش از ۱۷۰ میلیون نفر را در سراسر جهان آلوده کرده است (وامتهوین و پاندی، ۲۰۲۰).

تشخیص زودهنگام هپاتیت C برای درمان و پیش‌آگهی بسیار مهم است، در سال‌های اخیر،

تحقیقات زیادی برای توسعه سیستم‌های تشخیص زودهنگام هیپاتیت C با استفاده از الگوریتم‌های یادگیری ماشین سنتی انجام شده است (ممدوح و همکاران، ۲۰۲۲؛ هوان و نگوین، ۲۰۲۳؛ چن و همکاران، ۲۰۱۸؛ گاوه و همکاران، ۲۰۱۶). اگرچه مدل‌های موجود دقت نسبتاً خوبی دارند، اما هنوز اعتماد کامل پزشکان و بیماران را جلب نکرده‌اند.

الگوریتم‌های یادگیری عمیق قادر به تحلیل و پردازش حجم عظیمی از داده‌ها با دقت بسیار بالاتر نسبت به الگوریتم‌های سنتی هستند (گو و آلمتن، ۲۰۲۰). این الگوریتم‌ها می‌توانند الگوها و روابط پیچیده در داده‌های پزشکی را که ممکن است توسط انسان یا الگوریتم‌های سنتی شناسایی نشود، تشخیص دهند و در نتیجه دقت تشخیص و پیش‌بینی و وضعیت بیماران به‌طور قابل توجهی افزایش می‌یابد (ایکینز و پوهل، ۲۰۲۰). یکی از ویژگی‌های بارز الگوریتم‌های یادگیری عمیق، قابلیت خودآموزی و بهبود مستمر آن‌هاست. این الگوریتم‌ها با دریافت داده‌های جدید و به‌روزرسانی مدل‌های خود، به مرور زمان دقت بیشتری پیدا می‌کنند (علی‌زارگر و همکاران، ۲۰۲۳).

در دنیای پزشکی، داده‌های بسیاری از منابع مختلف مانند آزمایش‌های خون، تصاویر پزشکی و سوابق بیمار جمع‌آوری می‌شود و الگوریتم‌های یادگیری عمیق قادر به پردازش و تحلیل این داده‌های بزرگ و پیچیده به صورت کارآمد هستند (چن و وونگ، ۲۰۱۹). تحلیل دقیق داده‌های بالینی و بیوشیمیایی می‌تواند به تشخیص زودهنگام و پیش‌بینی بهتر کمک کند. این الگوریتم‌ها همچنین توانایی ارائه درمان‌های شخصی سازی شده را دارند (ممدوح و همکاران، ۲۰۲۲). با تحلیل دقیق داده‌های فردی بیماران، راهکارهای درمانی منحصر به فردی پیشنهاد می‌شود که با نیازها و شرایط خاص هر بیمار سازگار است و می‌تواند بهبود قابل توجهی در نتایج درمانی و کاهش عوارض جانبی داشته باشد (رسلی و همکاران، ۲۰۱۸).

لذا استفاده از الگوریتم‌های یادگیری عمیق در تشخیص و پیش‌بینی و وضعیت بیماران مبتلا به

هیپاتیت C از اهمیت و ضرورت ویژه‌ای برخوردار است. این الگوریتم‌ها با دقت و کارایی بالا، قابلیت خودآموزی، و توانایی پردازش داده‌های بزرگ و پیچیده، می‌توانند بهبود قابل توجهی در نتایج تشخیص و درمان این بیماری ایجاد کنند. با رفع چالش‌های موجود و بهره‌گیری از این فناوری‌ها، می‌توان گام بزرگی در جهت مقابله با هیپاتیت C و بهبود سلامت عمومی برداشت.

در رویکرد پیشنهادی پژوهشگر سعی کرده است تا با بکارگیری از الگوریتم یادگیری عمیق در زبان برنامه نویسی پایتون^۱، معایب الگوریتم‌های یادگیری ماشین سنتی را که پیش‌تر ذکر گردیده بود را بهبود دهد. لذا از الگوریتم یادگیری عمیق CNN و روش بالانس کردن داده‌ها در پیش پردازش، به عنوان یک نوآوری در پیش‌بینی و وضعیت بیماران مجموعه داده هیپاتیت C استفاده شده است. به این شکل که ابتدا مجموعه داده هیپاتیت C با استفاده از روش‌های پیش-پردازش از لحاظ ساختار کمی و محتوای کیفی بهبود یافته سپس با استفاده از الگوریتم یادگیری عمیق CNN به تشخیص وضعیت بیماران پرداخته می‌شود.

۱. مبانی نظری

در زمینه استفاده از الگوریتم‌های یادگیری ماشین و عمیق برای پیش‌بینی و تشخیص هیپاتیت C، چندین مطالعه قابل توجه انجام شده‌اند که از روش‌های مختلفی بهره برده‌اند، در ادامه برخی از آن‌ها تشریح داده شده‌اند.

(هوان و نگوین، ۲۰۲۳) از داده‌های تست خون روزانه برای تشخیص بیماری هیپاتیت C استفاده کرده‌اند. آن‌ها از شش الگوریتم ماشین بردار پشتیبان، K نزدیک‌ترین هم‌سایه، رگرسیون لجستیک، درخت تصمیم، XGBoost و شبکه عصبی مصنوعی برای یافتن مدل‌های مناسب که بتوانند بیماری هیپاتیت C را تشخیص دهند، استفاده کرده‌اند. در نهایت الگوریتم‌های SVM و XGBoost بهترین دقت (۹۶٫۸) را از میان سایر الگوریتم‌های داشتند.

^۱ Python

(ممدوح و همکاران، ۲۰۲۲) چهار مدل یادگیری ماشین شامل بیز ساده (NB)، RF، KNN و رگرسیون لجستیک را برای پیش‌بینی هپاتیت C با مجموعه داده‌ای شامل ۸۵۹ بیمار توسعه دادند. در نهایت مدل RF با دقت ۹۴.۸۸٪ به عنوان بهترین الگوریتم برای تشخیص بیماران هپاتیت C انتخاب شد.

(ما و همکاران، ۲۰۲۰) مدل‌های مختلف دسته‌بندی را برای پیش‌بینی بیماران هپاتیت C داده‌های جمع‌آوری شده از ۶۱۵ نفر ارائه دادند. در مطالعه آن‌ها، الگوریتم XGBoost با دقت ۹۱.۵۶٪ از سایر مدل‌ها بهتر عمل کرد.

(هاشم و همکاران، ۲۰۱۸) چندین رویکرد یادگیری ماشین را برای توسعه مدل‌های دسته‌بندی برای پیش‌بینی هپاتیت C مزمن پیشرفته با استفاده از داده‌های بالینی اعمال کردند. آن‌ها مدل‌های درخت تصمیم، رگرسیون خطی چند متغیره، بهینه‌سازی گروه ذرات و الگوریتم‌های ژنتیک را برای پیش‌بینی هپاتیت C مزمن پیشرفته توسعه دادند. برای ارزیابی عملکرد مدل‌های پیشنهادی، از مشخصه‌های عملکرد اخباری (ROC) استفاده شد. با دقت در بازه‌ای از ۶۶.۳٪ تا ۸۴.۴٪ و AUROC از ۰.۷۳ تا ۰.۷۶، این الگوریتم‌ها توانستند به‌طور موفقیت‌آمیز هپاتیت C مزمن پیشرفته را در بیماران مبتلا به کارسینوم هپاتوسلولی (HCC)، را پیش‌بینی کنند.

(علی‌زارگر و همکاران، ۲۰۲۳) یک مدل جدید برای تشخیص مرحله‌های مختلف هپاتیت C با استفاده از شبکه عصبی مصنوعی پس‌انتشار (ANN) ارائه داده است. مدل پیشنهادی شامل لایه ورودی، لایه پنهان و لایه خروجی است که با استفاده از داده‌های دریافتی از بیماران، وجود هپاتیت C را پیش‌بینی می‌کند. این مدل دقت بالایی را در پیش‌بینی مراحل مختلف بیماری نشان داده است.

(احمد و همکاران، ۲۰۲۰) در این مطالعه سه الگوریتم یادگیری ماشین را پیاده‌سازی نموده و

نشان دادند که KNN با دقت ۹۴.۴۰٪ بهترین عملکرد را دارد. (ناندپاتی و همکاران، ۲۰۲۰) دریافتند که برچسب‌های دو کلاسی در مقایسه با برچسب‌های چند کلاسی عملکرد بهتری دارند و با استفاده از مدل جنگل تصادفی (RF) به دقت ۵۴.۵۶٪ دست یافتند. (عبدالسلام و همکاران، ۲۰۱۹) آنها از مدل‌های چندگانه طبقه‌بندی استفاده کرده و به دقت‌هایی بین ۶۵.۶٪ تا ۶۸.۹٪ دست یافتند. (سیاف و همکاران، ۲۰۲۱) آنها چندین الگوریتم را ارزیابی کرده و دریافتند که شبکه‌های عصبی با دقت ۹۵.۱۲٪ بهترین عملکرد را دارند.

(شی و همکاران، ۲۰۲۰) چهار تکنیک یادگیری ماشین را استفاده کرده و دریافتند که روش درخت تصمیم با دقت ۹۳.۴۴٪ بهترین عملکرد را برای طبقه‌بندی و تشخیص هپاتیت C دارد. این مطالعات نشان می‌دهند که الگوریتم‌های مختلف یادگیری ماشین می‌توانند ابزارهای مؤثری برای تشخیص و پیش‌بینی هپاتیت C باشند. همچنین، استفاده از داده‌های آزمایش خون معمولی و ارزان قیمت می‌تواند به افزایش دقت پیش‌بینی کمک کند. هدف این مطالعات انتخاب بهترین الگوریتم‌ها برای پیش‌بینی هپاتیت C بر اساس داده‌های آزمایش خون معمولی و ارزان قیمت بوده است. در ادامه به بررسی و ارائه رویکرد پیشنهادی برای تشخیص وضعیت بیماران هپاتیت C با استفاده از الگوریتم CNN پرداخته شده است.

یادگیری عمیق به عنوان ابزاری قدرتمند در حال دگرگونی فرآیند کشف و توسعه دارو است. این فناوری با استفاده از شبکه‌های عصبی مصنوعی قادر به یادگیری از داده‌های حجیم و پیچیده و انجام وظایف مختلف با دقت بالا است. جدول ۱، برخی از ابزارها و الگوریتم‌های رایج یادگیری عمیق را نشان می‌دهد.

جدول ۱: الگوریتم‌های یادگیری عمیق در طراحی دارو (جنتایل و همکاران، ۲۰۲۰؛ کورتکوف و همکاران، ۲۰۱۷؛ ریفاوگلو و همکاران، ۲۰۱۹)

نام الگوریتم	شبکه‌های عصبی کانولوشنی (CNNs)	شبکه‌های عصبی بازگشتی (RNNs)	شبکه‌های عصبی تقویتی (RLNs)	یادگیری تقویتی عمیق (Deep RL)
کاربردها	✓ پیش‌بینی خواص مولکول‌ها ✓ طراحی مولکول‌های جدید ✓ پیش‌بینی پاسخ به درمان	✓ پیش‌بینی مسیرهای متابولیکی ✓ طراحی داروهای پروتئینی	✓ بهینه‌سازی فرآیند کشف و توسعه دارو ✓ طراحی مولکول‌های جدید با خواص مطلوب	✓ طراحی مولکول‌های جدید با خواص مطلوب ✓ بهینه‌سازی فرآیند تولید دارو
مزایا	✓ توانایی یادگیری الگوهای پیچیده از داده‌ها ✓ دقت بالا	✓ توانایی مدل‌سازی توالی‌های زمانی ✓ انعطاف‌پذیری بالا	✓ توانایی یادگیری از طریق آزمون و خطا ✓ انعطاف‌پذیری بالا	✓ ترکیب قدرت یادگیری عمیق و یادگیری تقویتی ✓ توانایی یادگیری از طریق آزمون و خطا
معایب	✓ نیاز به حجم زیادی از داده‌ها ✓ تفسیر نتایج دشوار	✓ آموزش دشوار ✓ حساسیت به نویز	✓ آموزش دشوار ✓ نیاز به زمان و منابع محاسباتی زیاد	✓ آموزش دشوار ✓ نیاز به زمان و منابع محاسباتی زیاد

در ادامه به معرفی الگوریتم یادگیری عمیق CNN و انواع پارامترهای ارزیابی الگوریتم‌های یادگیری عمیق در تشخیص بیماری‌ها پرداخته می‌شود، از این پارامترها در بخش ۳ برای ارزیابی رویکرد پیشنهادی استفاده شده است.

شبکه‌های عصبی پیچشی (CNN)

شبکه‌های عصبی پیچشی یا CNN (Convolutional Neural Networks) نقش مهمی در پیشرفت‌های اخیر در حوزه یادگیری عمیق و بینایی ماشین ایفا کرده‌اند (صالحی و همکاران، ۲۰۲۳). در ادامه، به توضیح مفاهیم اساسی و کاربردهای CNN پرداخته شده است.

ساختار و عملکرد CNN

لایه پیچشی (Convolutional Layer)

- هسته (Kernel) یا فیلتر (Filter) : یک ماتریس کوچک از وزن‌ها است که روی تصویر ورودی حرکت می‌کند (عملیات پیچش) و ویژگی‌های محلی مانند لبه‌ها، بافت‌ها یا الگوهای خاص را استخراج می‌کند (صالحی و همکاران، ۲۰۲۳).
- نقشه ویژگی (Feature Map) : خروجی حاصل از اعمال هسته‌ها بر روی تصویر ورودی که ویژگی‌های استخراج شده را نشان می‌دهد (ارکین و همکاران، ۲۰۲۳).

لایه تجمع (Pooling Layer)

- تجمع بیشینه (Max Pooling) : معمولاً استفاده می‌شود و حداکثر مقدار در هر ناحیه کوچک از نقشه ویژگی را انتخاب می‌کند. این کار باعث کاهش ابعاد داده و حفظ ویژگی‌های مهم می‌شود (شانگ و همکاران، ۲۰۲۴).
- تجمع میانگین (Average Pooling) : میانگین مقادیر در هر ناحیه کوچک را محاسبه می‌کند (رزق و همکاران، ۲۰۲۳).

لایه‌های تمام متصل (Fully Connected Layers)

- پس از چندین لایه پیچشی و تجمع، نقشه‌های ویژگی به یک بردار مسطح تبدیل می‌شوند که به لایه‌های تمام متصل متصل می‌شود. این لایه‌ها برای انجام طبقه‌بندی یا پیش‌بینی استفاده می‌شوند (صالحی و همکاران، ۲۰۲۳).

لایه خروجی (Output Layer)

- بسته به نوع مسئله (طبقه‌بندی، تشخیص شیء و غیره)، لایه خروجی شامل یک تابع فعال‌سازی مانند softmax برای طبقه‌بندی است که نتایج نهایی را ارائه می‌دهد (شانگ و همکاران، ۲۰۲۴).

مزایا و ویژگی‌های CNN

- **استخراج ویژگی خودکار:** برخلاف روش‌های سنتی که نیاز به استخراج دستی ویژگی‌ها داشتند، CNN ها به طور خودکار ویژگی‌های مناسب را از تصاویر استخراج می‌کنند (ارکین و همکاران، ۲۰۲۳).
- **محلی بودن ارتباطات:** فیلترهای پیچشی تنها بخش‌های محلی تصویر را بررسی می‌کنند که باعث کاهش تعداد وزن‌های قابل یادگیری و افزایش کارایی شبکه می‌شود (صالحی و همکاران، ۲۰۲۳).
- **مقیاس‌پذیری:** CNN ها می‌توانند با انواع مختلف داده‌های تصویری کار کنند و به راحتی با اندازه و مقیاس تصویر سازگار شوند (صالحی و همکاران، ۲۰۲۳).
- **کاهش ابعاد:** با استفاده از لایه‌های تجمع، ابعاد داده‌ها کاهش یافته و پردازش آسان‌تر می‌شود در حالی که ویژگی‌های مهم حفظ می‌شوند (ارکین و همکاران، ۲۰۲۳).

کاربردهای CNN

- **تشخیص اشیاء و چهره:** CNN ها به طور گسترده در تشخیص و شناسایی اشیاء و

- چهره‌ها در تصاویر و ویدئوها استفاده می‌شوند (صالحی و همکاران، ۲۰۲۳).
 - **تصویرسازی پزشکی** : در تشخیص بیماری‌ها از تصاویر پزشکی مانند MRI ، CT اسکن و تصاویر رادیولوژی کاربرد دارند (ارکین و همکاران، ۲۰۲۳).
 - **خودروهای خودران** : در تحلیل و درک تصاویر دوربین‌های نصب شده روی خودروها برای شناسایی موانع و علائم راهنمایی رانندگی (صالحی و همکاران، ۲۰۲۳).
 - **فیلترهای عکس و واقعیت افزوده** : در برنامه‌های موبایل برای اعمال فیلترهای واقعیت افزوده بر روی تصاویر و ویدئوها (ارکین و همکاران، ۲۰۲۳).
- شبکه‌های عصبی پیچشی (CNN) یکی از قدرتمندترین و پرکاربردترین الگوریتم‌های یادگیری عمیق هستند که با توانایی بالای خود در استخراج ویژگی‌های مهم از داده‌های تصویری، انقلابی در زمینه‌های مختلف از جمله بینایی ماشین، پزشکی و فناوری‌های هوشمند ایجاد کرده‌اند. این شبکه‌ها با ساختار خاص و لایه‌های متعدد خود، توانایی شناسایی و طبقه‌بندی دقیق تصاویر را دارند و در بسیاری از برنامه‌های کاربردی دنیای واقعی به کار گرفته می‌شوند.

پارامترهای ارزیابی

در ارزیابی مطالعات پیش‌بینی بیماری‌ها با استفاده از یادگیری عمیق، معمولاً از چندین پارامتر مختلف برای سنجش عملکرد مدل‌ها و کیفیت نتایج استفاده می‌شود. در زیر پارامترهای اصلی ارزیابی این نوع مطالعات به همراه توضیح و رابطه ریاضی ارائه شده است:

۱. معیار دقت (Accuracy): نسبت تعداد نمونه‌های درست تشخیص داده شده توسط مدل به کل تعداد نمونه‌ها. دقت به صورت رابطه ۱ محاسبه می‌شود (کرکاسکایت و همکاران، ۲۰۱۶).

$$\text{رابطه ۱: } \frac{TP + TN}{TP + TN + FP + FN}$$

که TP تعداد نمونه‌های مثبت درست، TN تعداد نمونه‌های منفی در ست، FP تعداد نمونه‌های مثبت اشتباه و FN تعداد نمونه‌های منفی اشتباه است.

۲. معیار فراخوانی (Recall): نسبت نمونه‌های مثبت درست پیش‌بینی شده به مجموع نمونه‌های مثبت واقعی است. فراخوانی به صورت رابطه ۲ محاسبه می‌شود (کرکاو سکایت و همکاران، ۲۰۱۶).

$$\text{رابطه ۲: } \frac{TP}{TP + FN}$$

۳. معیار واقعاً مثبت (Precision): نسبت نمونه‌های مثبت درست پیش‌بینی شده به مجموع نمونه‌های پیش‌بینی شده به عنوان مثبت است. معیار واقعاً مثبت به صورت رابطه ۳ محاسبه می‌شود (کرکاو سکایت و همکاران، ۲۰۱۶).

$$\text{رابطه ۳: } \frac{TP}{FP + TP}$$

۴. معیار F1 (F1 Score): یک میانگین هارمونیک از معیار فراخوانی و معیار واقعاً مثبت است و برای مواقعی که دسته‌بندی نمونه‌ها نیاز به توازن بین معیار فراخوانی و معیار واقعاً مثبت دارد، مورد استفاده قرار می‌گیرد. معیار F1 به صورت رابطه ۴ محاسبه می‌شود (کرکاو سکایت و همکاران، ۲۰۱۶).

$$\text{رابطه ۴: } \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

۵. منحنی ROC: یک رسم نمودار از نرخ واقعی مثبت (TPR) در مقابل نرخ اشتباه مثبت (FPR) برای مدل‌های طبقه‌بندی. منحنی ROC به عنوان یک معیار برای اندازه‌گیری عملکرد مدل‌های طبقه‌بندی استفاده می‌شود (ما و همکاران، ۲۰۲۰). خوب، بیاید یک مثال عددی برای استفاده از پارامترهای ارزیابی در طراحی دارو با استفاده از یادگیری عمیق را ببینیم. فرض کنید که ما از یک مدل یادگیری عمیق برای پیش‌بینی فعالیت دارویی استفاده کرده‌ایم. برای ارزیابی این مدل، ما دو کلاس داریم: فعال و غیرفعال. بر اساس خروجی مدل، داروها به عنوان فعال (مثبت) یا

غیرفعال (منفی) دسته‌بندی می‌شوند.

حال فرض کنید ما یک مدل آموزش داده‌ایم و از آن در مجموعه داده آزمون استفاده می‌کنیم. این مجموعه داده شامل ۱۰۰۰ نمونه است، از جمله ۴۰۰ نمونه فعال و ۶۰۰ نمونه غیرفعال است. حالا فرض کنید که ماتریس درهم‌ریختگی مدل ساخته شده به صورت جدول ۲ عمل کرده است:

جدول ۲: مقادیر پارامترهای ارزیابی

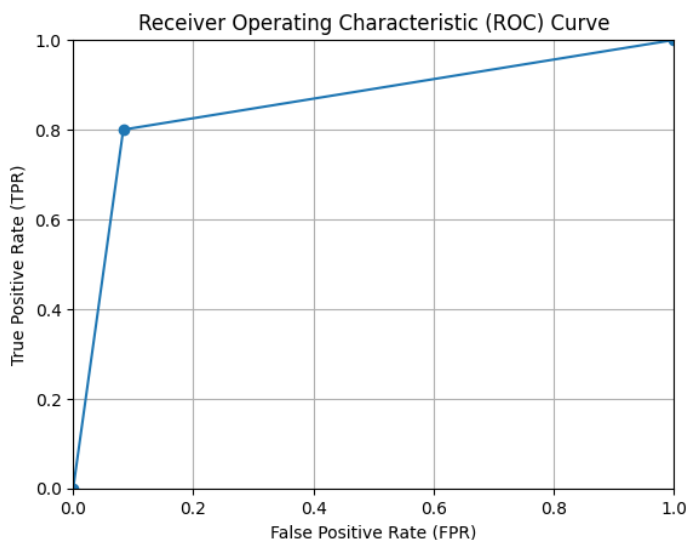
True Positives (TP)	False Positives (FP)	True Negatives (TN)	False Negatives (FN):
۳۲۰	۵۰	۵۵۰	۸۰

با استفاده از این اطلاعات، می‌توان پارامترهای ارزیابی را طبق جدول ۳ محاسبه کرد.

جدول ۳: نتایج پارامترهای ارزیابی

Accuracy	0.87
Recall	0.8
Precision	0.86
F1 Score	0.83

همچنین می‌توان برای ارزیابی مدل یادگیری عمیق ساخته شده از نمودار ROC نشان داده شده در شکل ۲ استفاده کرد.

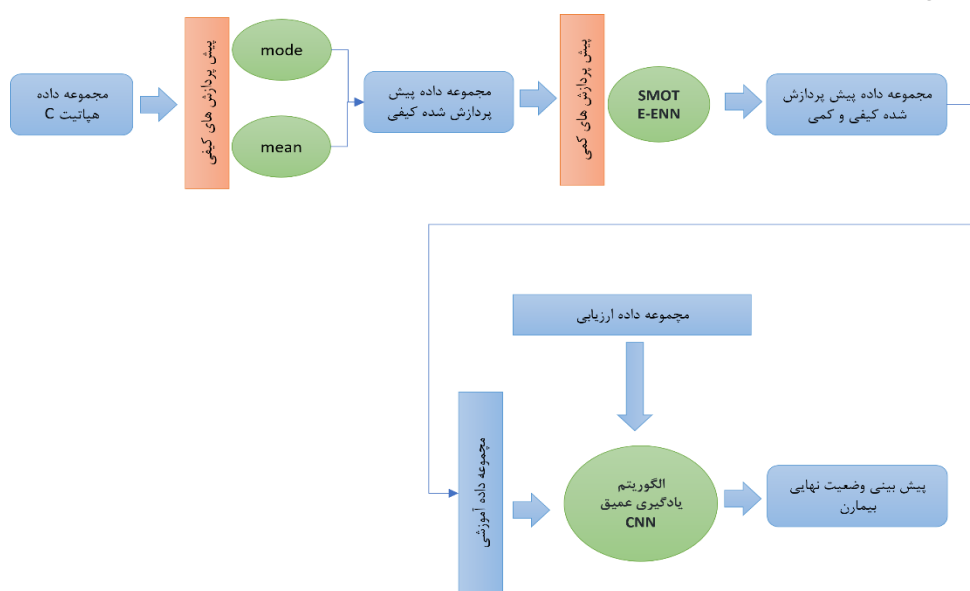


شکل ۲: نمودار ROC

نمودار ROC (Receiver Operating Characteristic) ارائه شده در شکل ۲ یکی از مهمترین ابزارهای ارزیابی در زمینه‌ی دسته‌بندی بر اساس مدل‌های آماری است. این نمودار به ما کمک می‌کند تا عملکرد مدل دسته‌بندی را در مخاطبی متغیر با توجه به تغییر حدآستانه به طور همزمان از دیدگاه دو معیار ارزیابی، یعنی True Positive Rate (TPR) و False Positive Rate (FPR) بررسی کنیم. در نمودار ROC محور افقی (X) مربوط به FPR و محور عمودی (Y) مربوط به TPR است. در نمودار مذکور، ما یک نقطه داریم که به مقدار FPR و TPR مربوط می‌شود. این نقطه نشان دهنده‌ی عملکرد مدل دسته‌بندی ما است. مساحت زیر نمودار (Area Under the Curve - AUC): مساحت زیر منحنی ROC (AUC) میزان کلی کیفیت مدل دسته‌بندی را نشان می‌دهد AUC بین ۰ و ۱ است، که یک AUC برابر با ۱ بهترین عملکرد را نشان می‌دهد و یک AUC برابر با ۰.۵ به عنوان عملکرد تصادفی مدل را نشان می‌دهد.

۲. روش‌شناسی تحقیق

در این پژوهش، یک رویکرد جامع برای تشخیص و پیش‌بینی وضعیت بیماران مبتلا به هپاتیت C ارائه شده است که شامل سه مرحله اصلی است (شکل ۳). در مرحله اول، مجموعه داده هپاتیت C از دانشگاه کارنگی ملون^۱ به عنوان داده ورودی استفاده شده است. سپس، فرایندهای پیش‌پردازش داده‌ها انجام شده است. این فرایندها شامل استفاده از روش‌های مد، میانگین و بالانس برای تکمیل مقادیر گم شده و متوازن کردن داده‌ها است. این اقدامات برای بهبود کیفیت داده‌ها و اطمینان از تعادل در مجموعه داده‌ها انجام شده است تا مدل بتواند با دقت بیشتری آموزش ببیند و نتایج معتبرتری ارائه دهد. در مرحله سوم، مدل یادگیری عمیق شبکه عصبی پیچشی (CNN) روی مجموعه داده پیش‌پردازش شده آموزش دیده و ارزیابی شده است. شبکه عصبی پیچشی به دلیل توانایی بالا در تشخیص الگوهای پیچیده در داده‌ها، انتخاب مناسبی برای این نوع پیش‌بینی‌ها است. پس از آموزش، مدل با استفاده از داده‌های آزمایشی ارزیابی می‌شود. در ادامه جزئیات هر یک از مراحل ذکر شده و نتایج نهایی ارزیابی مدل توسعه داده شده، در بخش ۳ ارائه شده است.



شکل ۳: مراحل رویکرد پیشنهادی

¹ Carnegie Mellon University (<https://archive.ics.uci.edu/dataset/46/hepatitis>)

۲.۱. مجموعه داده هیاتیت

مجموعه داده هیاتیت توسط دانشگاه کارنگی ملون به منظور توسعه و تست مدل‌های پیش‌بینی کننده جمع‌آوری و استفاده می‌شود (هیاتیت، ۱۹۸۸). این مجموعه داده بخصوص برای مسائل طبقه‌بندی و تحلیل پزشکی کاربرد دارد. (رسل و همکاران، ۲۰۱۸). مجموعه داده هیاتیت C دانشگاه کارنگی ملون معتبر است زیرا چندین عامل کلیدی اعتبار آن را تضمین می‌کنند. نخست، داده‌ها توسط محققان و متخصصان مجرب در دانشگاه کارنگی ملون جمع‌آوری و مدیریت می‌شوند، که خود یکی از معتبرترین موسسات آموزشی و پژوهشی در جهان است (هیاتیت، ۱۹۸۸). دوم، روش‌های علمی دقیق و استانداردهای بالا در جمع‌آوری و ثبت اطلاعات رعایت شده‌اند، که از صحت و دقت داده‌ها اطمینان می‌دهد (رسل و همکاران، ۲۰۱۸). سوم، این مجموعه داده به‌طور گسترده در پژوهش‌های علمی و مقالات معتبر استفاده و استناد شده است، که نشان‌دهنده پذیرش و تأیید جامعه علمی است (هیاتیت، ۱۹۸۸). نهایتاً، مجموعه داده‌ها به‌صورت شفاف و با مستندسازی کامل ارائه می‌شوند، که امکان بررسی و بازتولید نتایج را برای سایر محققان فراهم می‌سازد.

مجموعه داده هیاتیت شامل ۱۸ نوع از مشخصات و نتایج آزمایشات گرفته شده از ۱۵۵ بیمار مبتلا به هیاتیت C می‌باشد. ویژگی‌های موجود در این مجموعه داده از نظر پزشکی بسیار مهم بوده و می‌توانند برای تشخیص بیماری‌های کبدی یا فراهم آوردن پیشنهاداتی در مورد شدت بیماری و پیش‌بینی نتایج درمان استفاده شوند (احمد و همکاران، ۲۰۲۰).

این مجموعه داده اغلب برای آزمایش الگوریتم‌های طبقه‌بندی مثل درخت تصمیم، شبکه‌های عصبی، و ماشین‌های بردار پشتیبان استفاده می‌شود (رسل و همکاران، ۲۰۱۸). هدف از این آزمایش‌ها می‌تواند شامل پیش‌بینی پیشرفت بیماری، تعیین اثربخشی درمان‌ها، یا ارزیابی ریسک‌های پزشکی باشد (هیاتیت، ۱۹۸۸). دقت و کامل بودن مجموعه داده بسیار مهم است؛ برخی از ویژگی‌ها ممکن است داده‌های گمشده داشته باشند و در این صورت روش‌های مختلفی برای مقابله با داده‌های گمشده باید به کار گرفته شود. مجموعه داده هیاتیت UCI یک ابزار

ارزشمند برای جامعه علم داده است و امکان دسترسی به یک مجموعه داده واقعی را فراهم می‌کند که می‌تواند برای تقویت تحقیقات پزشکی و توسعه فناوری‌های جدید پزشکی استفاده شود (هپاتیت، ۱۹۸۸).

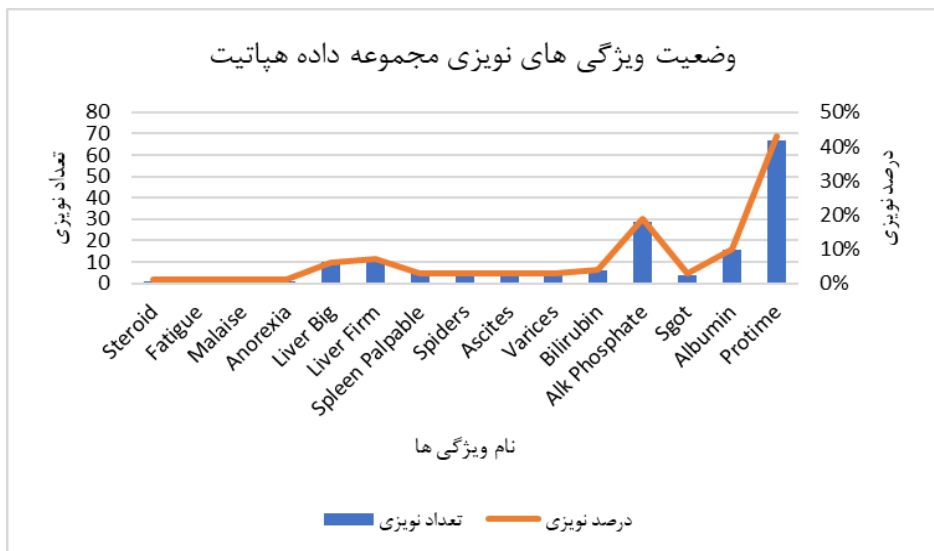
مجموعه داده هپاتیت C شامل ۱۹ ویژگی و یک برچسب کلاس است که نشان دهنده وضعیت بیمار است (مثلاً زنده یا فوت شده). ویژگی‌ها عبارتند از:

- سن (age): ارزش عددی با بازه سنی بیماران (شامل مقادیر عددی).
- جنس (sex): دسته‌بندی جنسیت بیماران (شامل مقادیر مرد یا زن).
- استروئید (steroid): استفاده‌ی بیمار از استروئیدها برای درمان (شامل مقادیر بله یا خیر).
- آنتی‌ویروس‌ها (antivirals): استفاده‌ی بیمار از داروهای ضد ویروس (شامل مقادیر بله یا خیر).
- خستگی (fatigue): وجود خستگی بیش از در بیماران (شامل مقادیر بله یا خیر).
- بدحالی (malaise): وجود ناخوشی، بی‌حالی و بدحالی در بیماران (بله یا خیر).
- درد کبد (malaise): وجود یا عدم وجود درد در ناحیه کبد (بله یا خیر).
- بی‌اشتهایی (anorexia): وجود یا عدم وجود بی‌اشتهایی در بیماران (بله یا خیر).
- کبد بزرگ (liver big): آیا کبد بزرگ‌تر از اندازه طبیعی است (بله یا خیر).
- کبد نرم (liver Firm): بافت کبدی به لمس نرم است یا خیر (بله یا خیر).
- طحال بزرگ (Spleen Palpable): آیا طحال بزرگ‌تر از حد طبیعی است (بله یا خیر).
- اسپایدرز (Spiders): وجود عنکبوت واریس که نشانه‌ی بیماری‌های کبدی است (بله یا خیر).

- آسیت (Ascites) : وجود مایع در حفره‌ی شکم که می‌تواند نشانه‌ی بیماری کبدی باشد (بله یا خیر).
 - واریس‌ها (Varices) : وجود واریس در معده یا مری (بله یا خیر).
 - بیلیروبین (Bilirubin) : سطح بیلیروبین در خون که میزان آن می‌تواند نشان‌دهنده‌ی آسیب کبدی باشد (بله یا خیر).
 - آلکالین فسفاتاز (Alk Phosphate) : آنزیمی که می‌تواند نشان‌دهنده‌ی آسیب کبدی باشد (شامل مقادیر عددی).
 - آمینوترانسفرازهای (sgot) : آنزیم دیگری که می‌تواند به ارزیابی آسیب کبدی کمک کند (شامل مقادیر عددی).
 - آلبومین (Albumin) : پروتئینی که در کبد تولید می‌شود و سطح آن می‌تواند نشان‌دهنده‌ی وضعیت عملکرد کبد باشد (شامل مقادیر عددی).
 - پروترومبین (protime) : زمان لازم برای انعقاد خون، که می‌تواند تحت تاثیر عملکرد کبد باشد (شامل مقادیر عددی).
 - هیستولوژی (Histology) : وجود تغییرات پاتولوژیک در بافت کبد که از طریق نمونه‌برداری از کبد (بیوپسی) تشخیص داده می‌شود (شامل مقادیر عددی).
- و در نهایت ویژگی وضعیت بیمار (Class) نیز به عنوان ویژگی هدف در مجموعه داده نشان دهنده زنده بودن و نبودن بیمار است. همه ویژگی‌ها براساس نتایج آزمایش و مشاهدات پزشکی جمع‌آوری شده‌اند. هنگام جمع‌آوری ویژگی‌ها از بیماران، اطلاعات مربوط به چند ویژگی جمع‌آوری نشده است که در بخش‌های بعدی با استفاده از روش‌های پیش‌پردازش این مورد بهبود داده شده است.

۲.۲. پیش پردازش داده ها

مقادیر برخی از ویژگی‌های بیماران هپاتیت در مجموعه داده ثبت نشده است که باعث وجود داده های نویزی و کاهش دقت نهایی مدل دسته‌بندی می‌گردد. شکل ۴ وضعیت ویژگی‌های نویزی در مجموعه داده هپاتیت C را نشان می‌دهد.



شکل ۴: وضعیت ویژگی‌های نویزی مجموعه داده هپاتیت C

با توجه به شکل ۴، حدود ۱۵ ویژگی دارای مقادیر نویزی می‌باشند که در این بین ویژگی‌های پروترومبین، البومین و آلکالین فسفاتاز دارای بیشترین مقادیر نویزی می‌باشند.

با استفاده از کتابخانه pandas مقادیر نویزی ۱۵ ویژگی مجموعه داده هپاتیت با توجه به نوع ساختار مقادیر آن‌ها بهبود پیدا کرد. لذا برای بهبود ویژگی‌های نویزی که دارای مقادیر اسمی بودند از روش مد و برای بهبود ویژگی‌های نویزی که دارای مقادیر عددی بودند از روش میانگین استفاده شد. روش‌های مد و میانگین دو تکنیک آماری هستند که می‌توانند برای کاهش

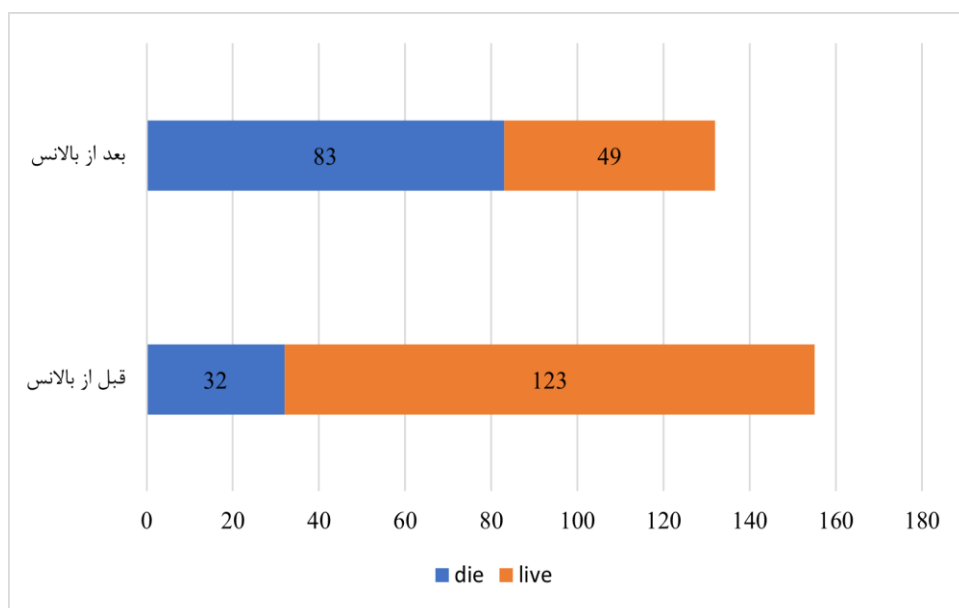
نویز موجود در ویژگی‌هایی که دارای تغییرات اتفاقی هستند، استفاده شوند. هر کدام از این روش‌ها به نحوی خاصی نویز را کاسته و نمایش بهتری از داده‌ها را ارائه می‌دهند. استفاده از روش‌های پیش‌پردازش مانند مد و میانگین در بسیاری از کارهای یادگیری ماشین و یادگیری عمیق (جنتایل و همکاران، ۲۰۲۰؛ کورتکوف و همکاران، ۲۰۱۷؛ ریفایوگلو و همکاران، ۲۰۱۹) به عنوان یک روش معمول پیش‌پردازش استفاده می‌شود.

استفاده از میانگین برای کاهش نویز در داده‌ها یکی از ساده‌ترین و مؤثرترین روش‌های پردازش سیگنال است (ناندیپاتی و همکاران، ۲۰۲۰). اگر داده‌های شما دارای نویز تصادفی باشند، میانگین‌گیری می‌تواند تأثیر نویز را کاهش دهد، زیرا نویزهای مثبت و منفی با یکدیگر خنثی می‌شوند (عبدالسلام و همکاران، ۲۰۱۹). برای اعمال این روش، شما می‌توانید مقادیر چندین مشاهده متوالی را بگیرید و میانگین آن‌ها را به عنوان مقدار جدید برای هر نقطه از داده‌ها در نظر بگیرید. این کار موجب می‌شود که نوسانات تصادفی کاهش یابند و داده‌ها صاف‌تر به نظر برسند (سیاف و همکاران، ۲۰۲۱). روش مد برای کاهش نویز در داده‌های دسته‌بندی شده یا گسسته کاربرد دارد. مد، مقداری است که بیشترین بار در داده‌ها ظاهر شده است (ناندیپاتی و همکاران، ۲۰۲۰). از این روش برای جایگزینی ویژگی‌های دارای نویز با مقداری که بیشترین تکرار را دارد، استفاده می‌شود. این کار به ویژه در مواردی که داده‌ها دچار خطاهای اندازه‌گیری یا تخمین‌های نادرست هستند، سودمند است (عبدالسلام و همکاران، ۲۰۱۹).

بالانس کردن مجموعه داده

یکی از معایب مجموعه داده هیپاتیت، توزیع نامتوازن کلاس‌های آن می‌باشد. از بین ۱۵۵ رکورد موجود در مجموعه داده هیپاتیت تنها ۳۲ رکورد آن دارای برچسب die بوده و مابقی دارای برچسب live می‌باشند. لذا مدل یادگیری عمیق ممکن است برای پیش‌بینی برچسب live

آموزش کاملتری ببیند. برای آموزش بالانس الگوریتم یادگیری عمیق باید مجموعه داده‌ای با توزیع متوازی از کلاس‌ها وجود داشته باشد. در رویکرد پیشنهادی علاوه بر الگوریتم یادگیری عمیق CNN، همچنین از روش ترکیب Over-Sampling و Under-Sampling به عنوان یک نوآوری، برای بالانس کردن کلاس‌های مجموعه داده هپاتیت استفاده شده است. شکل ۵ نتیجه قبل و بعد از فرآیند بالانس کردن مجموعه داده هپاتیت را با استفاده از روش ترکیبی نشان می‌دهد. همان طور که از شکل ۵ مشخص است بعد از فرآیند بالانس کردن مجموعه داده تعداد رکوردها با کلاس die به ۵۰ عدد افزایش یافته و تعداد رکوردها با کلاس live به ۸۳ عدد کاهش پیدا کرده است. لذا ۱۵۵ رکورد غیربالانس مجموعه داده اولیه هپاتیت به ۱۳۳ رکورد بالانس تبدیل می‌گردد.



شکل ۵: وضعیت رکوردهای مجموعه داده هپاتیت C قبل و بعد از بالانس

در ادامه از الگوریتم یادگیری عمیق CNN برای پیش بینی وضعیت نهایی بیماران هپاتیت C استفاده شده است. training set و test set به ترتیب ۸۰ و ۲۰ درصد از مجموعه داده هپاتیت

است که به ترتیب برای آموزش و ارزیابی مدل یادگیری عمیق CNN استفاده شده است. ماتریس در هم ریختگی با مقادیر پارامترهای TP، FP، FN و TN بدست آمده از نتایج پیش بینی مدل یادگیری عمیق CNN در شکل ۶ نشان داده شده است.

		برچسب های واقعی	
		Live	Die
برچسب های پیش بینی شده	Live	TP (38)	FP (1)
	Die	FN (0)	TN (7)

شکل ۶: ماتریس درهم ریختگی

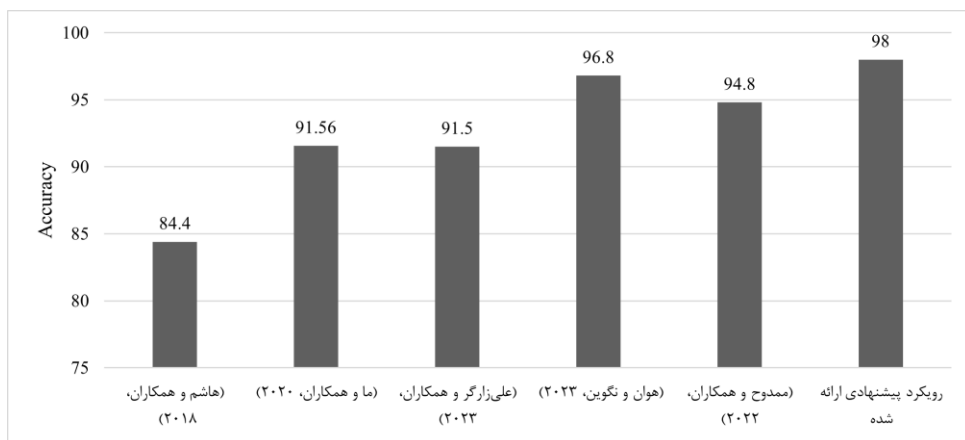
۳. بررسی یافته ها و نتایج

حال با بدست آوردن چهار پارامتر TP، FP، FN و TN در ماتریس درهم ریختگی، جدول ۴ مقادیر چهار پارامتر ارزیابی الگوریتم یادگیری عمیق CNN را قبل و بعد از فرآیند بالانس کردن مجموعه داده هپاتیت C نشان می دهد.

جدول ۴: مقادیر پارامترهای ارزیابی

	قبل از بالانس	بعد از بالانس
Accuracy	80%	98%
Recall	93%	100%
Precision	81%	97%
F1 Score	87%	98%

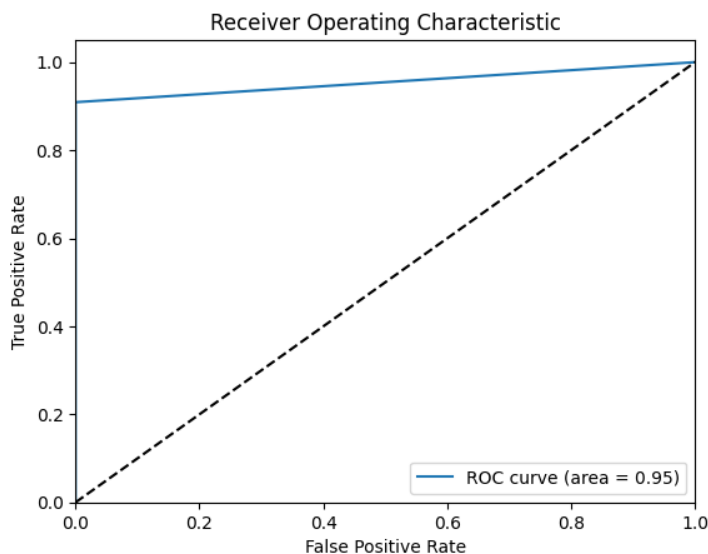
همانطور که از نتایج جدول ۴ مشخص است، فرآیند بالانس کردن مجموعه داده هیاتیت نقش مهمی در ارائه یک مدل یادگیری عمیق با دقت بالا دارد. شکل ۷ مقایسه نرخ دقت رویکرد پیشنهادی را با تعدادی از کارهای انجام شده بر روی مجموعه داده هیاتیت C نشان می‌دهد.



شکل ۷: مقایسه نرخ صحت رویکرد پیشنهادی با سایر کارهای انجام شده

شکل ۷ نشان می‌دهد که نرخ دقت رویکرد پیشنهادی نسبت به سایر کارها از برتری بیشتری برخوردار می‌باشد. لذا پزشکان با دقت و صحت بیشتری می‌توان به تشخیص وضعیت بیماران و ارائه تصمیم صحیح برای بهبود بیماران هیاتیت C پردازند.

شکل ۸ نمودار ROC الگوریتم یادگیری عمیق CNN را با توجه به مقادیر دو پارامتر TP و TN نشان می‌دهد. در نمودار ROC ارائه شده، منحنی ROC به طور قابل توجهی به سمت گوشه سمت چپ و بالا منحرف شده است. این نشان می‌دهد که مدل در تمایز بین نمونه‌های مثبت و منفی بسیار خوب عمل می‌کند. AUC برای این نمودار ۰.۹۵ است که نشان‌دهنده عملکرد بسیار خوب مدل است.



شکل ۸: نمودار ROC الگوریتم یادگیری عمیق CNN

۴. نتیجه گیری

استفاده از الگوریتم یادگیری عمیق CNN در تشخیص وضعیت بیماران هپاتیت C با دقت ۹۸٪ یک نقطه عطف مهم در توسعه فناوری پزشکی است. یادگیری عمیق، با بهره‌گیری از شبکه‌های عصبی عمیق، قادر به تحلیل دقیق داده‌های پزشکی هستند و الگوهای پیچیده در داده‌ها را شناسایی می‌کنند که به طور معمول برای انسان‌ها قابل تشخیص نیستند. این دقت بالا به پزشکان امکان می‌دهد تا با اطمینان بیشتری تشخیص دهند و برنامه‌ریزی درمانی را بهینه‌تر انجام دهند. علاوه بر این، استفاده از یادگیری عمیق بهبود مهمی در فرایندهای تشخیصی و مدیریت بیماری به وجود می‌آورد. این روش‌ها همچنین می‌توانند به پزشکان ایران و جهان به عنوان ابزاری پشتیبانی کننده در تصمیم‌گیری‌های بالینی ارائه شوند، از همین رو، اثربخشی درمان‌ها افزایش یافته و هزینه‌های بهداشتی کاهش می‌یابد. به طور کلی، این نگرش نه تنها به بهبود تشخیص

بیماری کمک می‌کند، بلکه به بهبود کیفیت مراقبت از بیماران و کاهش هزینه‌های مربوط به درمان نیز منجر می‌شود.

با توجه به استفاده از الگوریتم یادگیری عمیق CNN برای تشخیص بیماری هپاتیت C، می‌توان پیشنهادات زیر را برای کارهای آینده ارائه داد:

- گسترش مجموعه داده‌ها: استفاده از مجموعه داده‌های بزرگتر و متنوع‌تر می‌تواند دقت مدل را افزایش دهد و مدل را در برابر داده‌های ناشناخته مقاوم‌تر کند.
- ادغام با دیگر الگوریتم‌های یادگیری عمیق: ترکیب CNN با سایر الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی بازگشتی (RNN) یا مدل‌های توجه (attention models) می‌تواند قابلیت‌های پیش‌بینی مدل را بهبود بخشد.
- تحلیل چندوجهی داده‌ها: استفاده از داده‌های مختلف از جمله تصاویر پزشکی، نتایج آزمایش‌های بالینی و داده‌های ژنتیکی به صورت ترکیبی می‌تواند به مدل کمک کند تا با دقت بیشتری به تشخیص و پیش‌بینی بپردازد.
- پیش‌بینی پاسخ به درمان: توسعه مدل‌هایی که بتوانند پاسخ بیماران به درمان‌های مختلف را پیش‌بینی کنند، می‌تواند به شخصی‌سازی درمان‌ها و افزایش اثربخشی آن‌ها کمک کند.

فهرست منابع و مآخذ

1. Li, J., & Lu, H. (2020). Deep learning for drug discovery: An overview. *Drug Discovery Today*, 25(2), 308-317.
2. Goh, G. B., & Altman, R. B. (2020). Deep learning for molecular property prediction. *Journal of Chemical Information and Modeling*, 60(11), 5324-5333.
3. Ekins, S., & Puhl, A. C. (2020). Artificial intelligence for computer-aided drug discovery. *Nature Reviews Drug Discovery*, 19(11), 828-842.
4. Liu, Y., & Li, J. (2020). Applications of machine learning in precision oncology. *Frontiers in Oncology*, 10, 5824.
5. Mestres, J., & Gregori-Puigjané, E. (2020). Artificial intelligence for drug discovery: From target identification to clinical trials. *Nature Reviews Drug Discovery*, 19(11), 843-857.
6. Chan, H. C., & Wong, M. H. (2019). Deep learning for drug discovery and development. *Drug Discovery Today: Technologies*, 31, 19-26.
7. Stokes, J. M., & Yang, K. (2020). Deep learning for drug discovery. *Nature Reviews Drug Discovery*, 19(11), 810-827.
8. Brown, N., & Samanta, S. (2019). Applications of machine learning in drug discovery. *Current Opinion in Chemical Biology*, 51, 117-124.
9. Ching, T., Himmelstein, D. S., Weng, B. Q., & Beam, A. L. (2020). Opportunities and challenges for machine learning in clinical research. *Nature Medicine*, 26(11), 1591-1598.
10. Wang, J., Zhang, J., Lao, Z., & Liu, Y. (2020). Recent advances in machine learning for drug discovery. *Current Opinion in Structural Biology*, 62, 117-124.
11. Sachdeva, R. K., Bathla, P., Rani, P., Solanki, V., & Ahuja, R. (2023). A systematic method for diagnosis of hepatitis disease using machine learning. *Innovations in Systems and Software Engineering*, 19(1), 71-80.
12. Alizargar, A., Chang, Y. L., & Tan, T. H. (2023). Performance comparison of machine learning approaches on Hepatitis C prediction employing data mining techniques. *Bioengineering*, 10(4), 481.
13. Rosly, R., Makhtar, M., Awang, M. K., Awang, M. I., & Rahman, M. N. A. (2018). Analyzing performance of classifiers for medical datasets. *International Journal of Engineering & Technology*, 7(2.15), 136-138.
14. Mamdouh, H.; Shams, M.; Abd El-Hafeez, T. Hepatitis C Virus Prediction Based on Machine Learning Framework: A Real-World Case Study in Egypt. *Knowl. Inf. Syst.* 2022, 65, 2595–2617.
15. Huynh, P. H., & Nguyen, V. H. (2023). A Novel Ensemble of Support Vector Machines for Improving Medical Data Classification. *Engineering Innovations*,

- 4, 47-66.
16. Vamathevan, J., & Pandey, A. (2020). Deep learning in drug discovery: Applications and challenges. *Drug Discovery Today*, 25(2), 318-327.
17. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug discovery today*, 23(6), 1241-1250.
18. Gawehn, E., Hiss, J. A., & Schneider, G. (2016). Deep learning in drug discovery. *Molecular informatics*, 35(1), 3-14
19. Gentile, F., Agrawal, V., Hsing, M., Ton, A. T., Ban, F., Norinder, U., ... & Cherkasov, A. (2020). Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS central science*, 6(6), 939-949.
20. Korotcov, A., Tkachenko, V., Russo, D. P., & Ekins, S. (2017). Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Molecular pharmaceutics*, 14(12), 4462-4475.
21. Rifaioglu, A. S., Atas, H., Martin, M. J., Cetin-Atalay, R., Atalay, V., & Doğan, T. (2019). Recent applications of deep learning and machine intelligence on in silico drug discovery: methods,
22. Krakauskaitė, S., Petkus, V., Bartusis, L., Zakelis, R., Chomskis, R., Preiksaitis, A., ... & Rastenyte, D. (2016). Accuracy, precision, sensitivity, and specificity of noninvasive ICP absolute value measurements. *Intracranial Pressure and Brain Monitoring XV*, 317-321.
23. Ma, L.; Yang, Y.; Ge, X.; Wan, Y.; Sang, X. Prediction of Disease Progression of Chronic Hepatitis C Based on XGBoost Algorithm. In *Proceedings of the 2020 International Conference on Robots & Intelligent System (ICRIS)*, Sanya, China, 7–8 November 2020; pp. 598–601
24. Hashem, S.; Esmat, G.; Elakel, W.; Habashy, S.; Raouf, S.A.; Elhefnawi, M.; Eladawy, M.I.; ElHefnawi, M. Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2018, 15, 861–868.
25. Hepatitis. (1988). UCI Machine Learning Repository. <https://doi.org/10.24432/C5Q59J>.
26. Ahammed, K.; Satu, M.S.; Khan, M.I.; Whaiduzzaman, M. Predicting Infectious State of Hepatitis C Virus Affected Patient's Applying Machine Learning Methods. In *Proceedings of the 2020 IEEE Region 10 Symposium (TENSYP)*, Dhaka, Bangladesh, 5–7 June 2020; pp. 1371–1374
27. Nandipati, S.C.; XinYing, C.; Wah, K.K. Hepatitis C Virus (HCV) Prediction by Machine Learning Techniques. *Appl. Model. Simul.* 2020, 4, 89–100.
28. Abd El-Salam, S.M.; Ezz, M.M.; Hashem, S.; Elakel, W.; Salama, R.; ElMakhzangy, H.; ElHefnawi, M. Performance of Machine Learning Approaches on Prediction of Esophageal Varices for Egyptian Chronic Hepatitis C Patients. *Inform. Med. Unlocked* 2019, 17, 100267.
29. Syafa'ah, L.; Zulfatman, Z.; Pakaya, I.; Lestandy, M. Comparison of Machine

- Learning Classification Methods in Hepatitis C Virus. *J. Online Inform.* 2021, 6, 73
30. Shi, L.; Wei, L.; Tao, Y.; Oleiwi, A. Development of Diagnostic Decision Making For Chronic Hepatitis C Virus Patients By Various Supervised Predictive Model. *J. Adv. Res. Dyn. Control Syst.* 2020, 12,
 31. Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, 15(7), 5930.
 32. Arkin, E., Yadikar, N., Xu, X., Aysa, A., & Ubul, K. (2023). A survey: object detection methods from CNN to transformer. *Multimedia Tools and Applications*, 82(14), 21353-21383.
 33. Rizk, F. H., Arkhstan, S., Zaki, A. M., Kandel, M. A., & Towfek, S. K. (2023). Integrated CNN and Waterwheel Plant Algorithm for Enhanced Global Traffic Detection. *Journal of Artificial Intelligence and Metaheuristics*, 6(2), 36-6.
 34. Shang, S., Shan, Z., Liu, G., Wang, L., Wang, X., Zhang, Z., & Zhang, J. (2024, March). Resdiff: Combining cnn and diffusion model for image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 8, pp. 8975-8983).