۱- دکتر امیرحسین بحرالعلومیان؛
دکتری تخصصی مهندسی پزشکی
۲- مهندس منیره توکلی (کارشناس ارشد مهندسی پزشکی)

کاربرد هوش مصنوعی در تشخیص های آزمایشگاهی – بخش نهم

در ادامه، پژوهشگران COVID-19 و پزشکان به بررسی تکنیک‌های یادگیری پیشرفته هوش مصنوعی پرداختند تا گزینه‌های جایگزینی برای پیش‌بینی، مدیریت، و نظارت بر COVID-19 پیدا کنند و راه‌های قابل تعمیم و مقرون به صرفه‌ای برای مقابله با پاندمی شناسایی کنند. این تلاش‌ها به دنبال بهبود فرآیندهای تشخیصی و درمانی و همچنین بهینه‌سازی منابع در شرایط بحرانی است.

در مورد تشخیص، چندین مطالعه در اواسط ماه آوریل ۲۰۲۰ با بررسی آزمایش‌های خون روتین و استفاده از داده‌های اختصاصی (مرکز واحد) انجام شد که به خاطر اندازه کوچک نمونه ها (n<1,000) عمدتا بدون اعتبارسنجی خارجی ارائه شده اند.

جوشی و همکاران روشی جالب را ارائه دادند که در آن تشخیص بیماری با استفاده از رگرسیون لجستیک منظم‌شده L2 مدل‌سازی شده است. این مدل تنها با استفاده از سطوح هماتوکریت، نوتروفیل‌ها و لنفوسیت‌ها آموزش داده شده و موفق به دستیابی به اعتبارسنجی داخلی (AUC=0.78) شده است که با ارزیابی انجام شده در چهار سایت مختلف (میانگین AUC=0.77) سازگار بود (۲۰۲۰).

بریناتی و همکاران از رگرسیون لجستیک استفاده کردند، عملکرد طبقه‌بندی آن را با یک طبقه‌بند جنگل تصادفی مقایسه کردند و از ۱۴ ویژگی برای این کار بهره بردند (۲۰۲۰). نتایج در مجموعه آزمایش داخلی بسیار مشابه بودند، اما درخت تصمیم توانست درک بهتری ارائه دهد و نشان داد که AST (<25.4) و لنفوسیت‌ها (<۱,۳) به عنوان پیش‌بینی‌کننده‌های اصلی منفی بودن COVID-19 شناخته می‌شوند.

آلوز و همکاران از یک طبقه‌بند جنگل تصادفی استفاده کردند و عملکرد طبقه‌بندی آن را با پنج الگوریتم دیگر مقایسه کردند، این مجموعه بهترین طبقه‌بندی داخلی را با AUC=0.87 به دست آورد (۲۰۲۱). یک درخت تصمیم برای توضیح مدل استفاده شد و با استفاده از نمودارهای معیار، می توان ارتباط بین پارامترهای خونی انتخاب شده را به صورت چشمی تفسیر کرد.

بانرجی و همکاران یک شبکه عصبی مصنوعی طراحی کردند و آن را با یک جنگل تصادفی و یک مدل خطی تعمیم‌یافته منظم‌شده با لاسو-الاستیک‌نت (که معادل رگرسیون لجستیک است) مقایسه کردند (۲۰۲۰). این شبکه در افراد جامعه (n=619) و بیماران در بخش عادی بیمارستان (n=69) آزمایش شد. در حالی که شبکه عصبی مصنوعی (ANN) و جنگل تصادفی (RF) بهترین معیارها را برای بیماران بستری و غیربستری ارائه دادند، مدل glmnet یک الگوی کاهشی در مونوسیت‌ها، لکوسیت‌ها، ائوزینوفیل‌ها و پلاکت‌ها شناسایی کرد که در رگرسیون لجستیک به کار گرفته شد و موفق به دستیابی به AUC برابر با ۰/۸۵ شد. مدل تجمیعی طراحی شده توسط آبا یومی-آلی و همکاران تحت یک مجموعه داده کوچک (n=279) ساخته شد که ۱۶ ویژگی را به عنوان ورودی در نظر می گیرد. مقایسه‌هایی بین ۱۵ طبقه‌بند انجام شد که در آن‌ها مدل ExtraTrees با AUC=0.99 و مدل AdaBoost با AUC=0.98 از سایر مدل‌ها بهتر عمل کردند.

وو و همکاران نیز با استفاده از یک روش انتخاب مجموعه پویا و یک مجموعه داده کمی بزرگ‌تر (n=603) مقادیر طبقه‌بندی داخلی مشابهی را به دست آوردند. این روش ابتدا با تکنیک‌های تعادل داده‌ها رویارو شد و سپس با یک خوشه‌بندی ترکیبی همراه با یک طبقه‌بند بگینگ پسین مدل‌سازی شد (۲۰۲۱). نویسندگان با استفاده از رویکردهایبرید، نتایج بهتری نسبت به استفاده از رویکرد بگینگ به دست آوردند که در تقسیم‌بندی‌های ۷۰:۳۰ و ۶۰:۴۰ و با استفاده از ارزیابی متقابل ۵-تایی آزمایش شد.

برخلاف مطالعات نظارت‌شده قبلی، سوزا و همکاران یک رویکرد گزارش کردند براساس خوشه‌بندی غیرنظارت‌شده مبتنی

بر نقشه‌های سازمان‌دهی، که با مدل LDA توانست بیماران مبتلا به COVID-19 را با قدرت تفکیک ۸۳٪ شناسایی کند (۲۰۲۱). این رویکرد خوشه‌بندی بر روی ۵۹۹ مورد انجام شد، از این تعداد تنها ۸۱ مورد مثبت COVID-19 بودند. این روش، WBC لکوسیت‌ها، BA بازوفیل‌ها، (EO ائوزینوفیل‌ها) و RDW دامنه توزیع گلبول‌های قرمز را به عنوان ویژگی‌هایی با تأثیر قوی بر عملکرد خوشه‌بندی شناسایی کرد، اما دامنه ویژگی‌ها در خروجی پیش‌بینی ابهام داشت. این مطالعات با استفاده از الگوریتم‌های پیچیده‌تر یادگیری ماشین دقت نتایج را بهبود بخشید، مطالعاتی که دارای اندازه نمونه بزرگ‌تری (n>1,000) بود، نشان داد که با اضافه شدن ویژگی‌های خونی، معیارهای طبقه‌بندی نیز افزایش مشابهی دارند.

تسچولیتس و همکاران در یک گروه ۱/۵۳۷ نفره از شرکت‌کنندگان، موفق به دستیابی به AUC متوسط ۰/۷۴ و ارزش پیش‌بینی منفی ۹۸٪ شدند که با نتایج قبلی استفاده از جنگل تصادفی همخوانی داشت (۲۰۲۱).

کابیتزا و همکاران یک روش نوآورانه برای ارزیابی قابلیت اطمینان مدل‌های طبقه‌بندی در شرایط اعتبارسنجی خارجی ارائه دادند. این روش بر دو معیار اصلی متمرکز است: ۱. تعداد (Cardinality) ۲. شباهت (Similarity). با توجه به داده‌های مربوط به مشخصات جمعیتی و شمارش کامل سلول‌های خونی، از ماشین بردار پشتیبانی (SVM) با هسته RBF برای هشت مجموعه داده خارجی مختلف استفاده شد. مقدار AUC این مدل در این مجموعه‌ها به ترتیب برابر با ۰/۶۶، ۰/۷۵، ۰/۸۰، ۰/۸۳، ۰/۸۷، ۰/۸۹، ۰/۹۷ و ۰/۹۸ بود. همچنین، مقادیر شباهت (براساس درجه تطابق) به ترتیب برابر با ۰/۳۱۵، ۰/۳۴۱، ۰/۴۴۵، ۰/۴۳۹ و ۰/۴۴۷، ۰/۳۲۳، ۰/۴۴۴، ۰/۳۴۸ بودند. بابایی و همکاران عملکرد ۱۲ الگوریتم یادگیری ماشین را در سه مجموعه داده مختلف مقایسه کردند. در سومین مجموعه داده، مقایسه عملکرد تمامی الگوریتم‌ها نشان داد که شبکه‌های عصبی عمیق (DNN) بالاترین معیارهای طبقه‌بندی را داشته‌اند (۲۰۲۲). جالب این است که مطالعات قبلی برناتی و همکاران (۲۰۲۰) و کابیتزا و همکاران (۲۰۲۱) نیز با شبکه‌های عصبی عمیق (DNN) مقایسه شدند که در مجموعه داده اول (AUC=0.92 در مقابل AUC=0.84، از برناتی و همکاران) و در مجموعه داده دوم (AUC=0.93 در مقابل AUC=0.84، از کابیتزا و همکاران) عملکرد بهتری داشتند. این نتایج نشان می‌دهد که شبکه‌های عصبی عمیق به عنوان یک رویکرد امیدوارکننده برای تشخیص COVID-19 مطرح هستند.

پلانت و همکاران از یک گروه بزرگ متشکل از ۶۶ بیمارستان برای انجام اعتبارسنجی داخلی و خارجی یک درخت تقویت شده با گرادیان شدید (Extreme Gradient Boosting) استفاده کردند که بر اساس ۱۵ ویژگی طراحی شده بود. اعتبارسنجی خارجی که در ۲۳ بیمارستان مختلف انجام شد، به تأیید روش مورد استفاده با AUC=0.91 کمک کرد و این امکان را فراهم کرد که به‌طور عمیق‌تری بهترین نمره برش (نمره‌ای که برای تشخیص مثبت یا منفی بیماری استفاده می‌شود) را درک کنیم. این بررسی به‌گونه‌ای انجام شد که تأثیر شیوع بیماری (در سه سطح ۱٪، ۱۰٪ و ۲۰٪) بر این نمره برش مستقل از هم مورد مطالعه قرار گرفت.

کمپانیجر و همکاران شش الگوریتم را در دو مکان مختلف (برگامو با ۲۴۵ شرکت‌کننده و دزیو با ۳۳۷ شرکت‌کننده) اعتبارسنجی کردند که در آن ۴۲٪ و ۴۸٪ مورد مثبت COVID-19 وجود داشت (۲۰۲۱). مدل‌ها همواره AUC بالاتری از ۹۳٪ را به دست آوردند، ماشین بردار پشتیبانی (SVM) بهترین نتایج را در هر دو مجموعه داده خارجی کسب کرد. نمودارهای ویولن (Violin plots) پارامترهای کلیدی شمارش کامل سلول‌های خونی (CBC)، نشان می دهند که گروه‌های آموزشی و اعتبارسنجی شباهت زیادی دارند. این شباهت‌ها شامل تعداد گلبول‌های سفید، نوتروفیل‌ها، لنفوسیت‌ها، گلبول‌های قرمز، تعداد پلاکت‌ها و سن بیماران بود که ثبات عملکرد مدل را نشان می دهد.

چادگا و همکاران از روش‌های مشابهی برای دو مجموعه داده عمومی استفاده کردند:

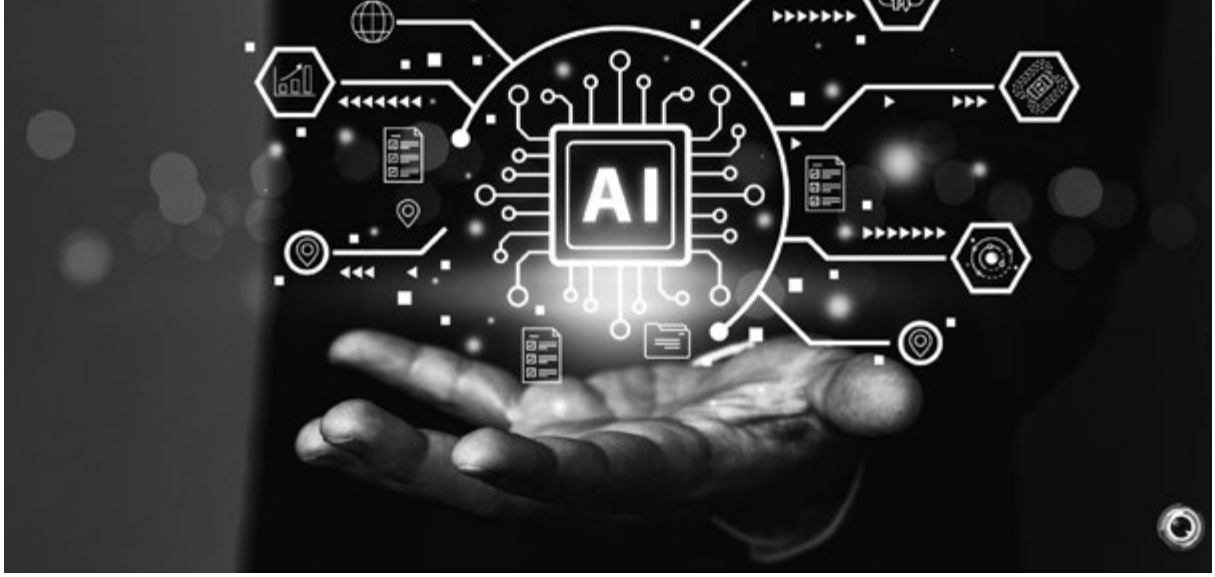
۱- بیمارستان آلبرت اینشتین در برزیل: در این مطالعه از ۵/۶۴۴ نمونه استفاده شد و الگوریتم جنگل تصادفی AUC(RF) برابر با ۰/۸۰ را به دست آورد.

۲- بیمارستان دکتر تی ام ای پای در هند: در این مطالعه ۱/۱۶۹ نمونه استفاده شد و الگوریتم جنگل تصادفی AUC بسیار بالاتری برابر با ۰.۹۹ را نشان داد.

هر دو مطالعه از روش SMOTE برای مقابله با داده‌های نامتوازن استفاده کردند. اما مطالعه دوم از روش‌های قابل تفسیر نیز برای توضیح نحوه تأثیر ویژگی‌ها بر تصمیم نهایی استفاده کرد. این کار باعث بهبود قابل توجه در معیارهای عملکرد شد (به ویژه با مقایسه الگوریتم جنگل تصادفی یکسان در دو مطالعه). با این حال، هیچ کدام از این مطالعات اعتبارسنجی خارجی انجام نداده‌اند.

بنیتو-لئون و همکاران از یک مدل خوشه‌بندی بدون نظارت (X-means) برای تشخیص شدت بیماری COVID-19 استفاده کردند. این مدل توانست بیماران مثبت را به سه دسته تقسیم کند: بیماران بستری در بخش مراقبت‌های ویژه، بیماران بستری عادی و بیماران غیر بستری (۲۰۲۱). بر اساس شاخص دیوید بولدین، که هر چه مقدار آن کمتر باشد، توزیع خوشه‌ها بهتر است (یعنی فاصله بین خوشه‌ها بیشتر و فاصله درون خوشه‌ها کمتر است)، الگوریتم سه خوشه را شناسایی کرد. مقدار فاصله منهتن برای این خوشه‌ها برابر با ۰/۷۰۱ بود. ویژگی‌های مرتبط برای تمایز بین خوشه‌ها با استفاده از مقادیر p و اندازه اثر ارزیابی شده‌اند.

فامیگلینی و همکاران از یک روش طبقه‌بندی نظارت شده



برای پیش‌بینی پذیرش بیماران COVID-19 در بخش مراقبت‌های ویژه (ICU) استفاده کردند. در این مطالعه، ۱۰۰۴ بیمار COVID-19 مورد بررسی قرار گرفتند که تنها ۱۸/۳٪ از آن‌ها به ICU منتقل شدند. این موضوع نشان‌دهنده عدم تعادل در داده‌ها است (۲۰۲۲). مدیریت داده‌ها (شامل پر کردن داده‌های گم‌شده و ارزیابی سوگیری) و انتخاب مدل مناسب، باعث بهبود قابل توجه در معیارهای عملکرد مدل شد:

۱ - امتیاز AUC (شاخص دقت طبقه‌بندی) به ۰/۸۵ رسید، که نشان‌دهنده عملکرد خوب مدل در تشخیص بیماران نیازمند بستری در ICU است.

۲ - امتیاز بریر (شاخص کالیبراسیون) به ۰/۱۴۴ کاهش یافت، که نشان می‌دهد پیش‌بینی‌های مدل با واقعیت نزدیک‌تر شده است.

۳ - نفع خالص استاندارد شده (شاخص کاربرد بالینی) به ۰/۶۹ رسید، که بیانگر افزایش سودمندی بالینی مدل است.

همچنین، مدل نشان داد که سطح NLR (نسبت نوتروفیل به لنفوسیت) از اهمیت بالایی در پیش‌بینی نیاز به بستری در ICU برخوردار است، که با یافته‌های قبلی در این زمینه همخوانی دارد.

لو و همکاران نیز این نتیجه را مورد بررسی قرار دادند و ۶۷ بیمار با شدت خفیف و ۱۲۹ بیمار با شدت بالا را مطالعه کردند. آن‌ها از یک سیستم هیبریدی مبتنی بر تصمیم‌گیری چندمعیاره (MCDM) استفاده کردند که شامل ترکیبی از الگوریتم TOPSIS (روش ترتیب ترجیح بر اساس شباهت به راه‌حل ایده‌آل) و یک طبقه‌بند نایو بیس بود.

روش TOPSIS پیش‌پردازش و رتبه‌بندی ویژگی‌ها را انجام می‌دهد، در حالی که نایو بیس (NB) انتخاب ویژگی‌ها را انجام می‌دهد. با وجود اینکه این روش موفق به کسب AUC بالاتری برابر با ۰/۹۳ شد، اما اندازه نمونه کوچک بود و شامل اعتبارسنجی خارجی نمی‌شد.

موری و همکاران مدلی برای پیش‌بینی پیش آگهی بیماران COVID-19 توسعه دادند. آن‌ها از یک مدل رگرسیون لجستیک قابل تفسیر استفاده کردند که بر اساس داده‌های ۹۲۱ بیمار بستری ساخته شده بود. از این تعداد، ۱۲۰ بیمار فوت کردند (با شیوع ۱۳٪) (۲۰۲۱). مدل توانایی بالایی در تفکیک بیماران بر اساس سطوح هموگلوبین، پلاکت‌ها، نوتروفیل‌ها، اوره، پروتئین C-reactive و سدیم داشت (AUC=0.87). با این حال، در اعتبارسنجی خارجی بعدی که بر روی جمعیتی با شیوع ۲۲/۶٪ انجام شد، این مقدار به (AUC=0.81) کاهش یافت.

فرناندز و همکاران مدل تشخیصی را برای مرگ و میر، ونتیلیشن مکانیکی تهاجمی و بستری در بخش مراقبت‌های ویژه (الگوریتم‌های چندمنظوره) گسترش دادند (۲۰۲۱). با استفاده از تعداد کمتری ویژگی (سن، نسبت لنفوسیت به پروتئین C-reactive، پروتئین C-reactive و نتایج مقیاس برادن)، نویسندگان به این نتیجه رسیدند که می‌توان هر یک از پیامدهای مورد بررسی (بستری در ICU، نیاز به ونتیلیشن مکانیکی تهاجمی یا مرگ و میر) را با استفاده از داده‌های مربوط به سایر پیامدها پیش‌بینی کرد. در تمام این مدل‌ها، مقدار AUC همواره بالاتر از ۹۱٪ بود که نشان‌دهنده دقت بالای آن‌ها در پیش‌بینی پیامدها است.

در مطالعه کارتیکایان و همکاران، XGBoost بر روی مجموعه داده‌ای متشکل از ۱۷۰ بیمار فوت شده و ۲۰۰ بیمار بهبود یافته اعمال شد تا اهمیت ویژگی‌ها و شبکه عصبی برای انتخاب ویژگی‌ها تعیین شود، و در نتیجه عملکرد پیش‌بینی بهتری به دست آمد. XGBoost یک الگوریتم درخت تصمیم بهبود یافته است که به طور موثری می‌تواند ویژگی‌های مهم را شناسایی کند. سپس شبکه عصبی برای انتخاب ویژگی‌های مناسب‌تر از میان ویژگی‌های مهم شناسایی شده توسط XGBoost استفاده شد. این رویکرد ترکیبی منجر به بهبود دقت پیش‌بینی مرگ و میر بیماران COVID-19 شد. ویژگی‌هایی که انتخاب شده بودند، توانستند تعداد روزهایی که تا وقوع باقی مانده را پیش‌بینی کنند. نتایج نشان داد که دقت این پیش‌بینی‌ها همیشه بالای ۹۰ درصد بوده است. این مدل‌ها با استفاده از داده‌هایی که تا ۱۲ روز قبل از وقوع رویداد جمع‌آوری شده بودند، آموزش دیده‌اند و شامل اطلاعاتی از روزهای نزدیک به وقوع رویداد نیستند (این موضوع به مورد ۲ اشاره دارد). همچنین نویسندگان الگوهای خونی مرتبط با پیش‌بینی مرگ و میر را نشان دادند، مانند مقادیر بالای LDH، hs-CRP و نوتروفیل‌ها و مقادیر پایین ائوزینوفیل‌ها که با ادبیات قبلی همخوانی دارد.

در شماره های بعدی به بررسی چالش های موجود در این حوزه خواهیم پرداخت.