

Personalization of the Radiology Residency Training Process Using Deep Learning and Interactive Extraction of Diagnostic Error Patterns

Seyed Ali Kianmehr^{1,2}, Mahdi Hashemzadeh^{1,2*}

¹ Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran

² Artificial Intelligence and Machine Learning Research Laboratory, Azarbaijan Shahid Madani University, Tabriz, Iran

Received: 20 January 2025, Revised: 30 August 2025, Accepted: 22 October 2025

Paper type: Research

Abstract

Radiological imaging is a fundamental diagnostic tool widely utilized by physicians for the diagnosis and monitoring of various medical conditions. Accurate interpretation of these images demands training of specialized residents and medical experts to ensure high diagnostic precision. This study presents an intelligent system that leverages machine learning to provide a personalized radiology residency training program through the interactive extraction of false-positive and false-negative diagnostic error patterns. The proposed approach is designed as an online personalized training platform that employs an iterative process to evaluate residents through adaptive tests generated by a question-selection algorithm based on each resident's performance, with the aim of extracting diagnostic error models of them independently. The proposed method employs a deep learning-based model to evaluate the difficulty level of educational radiographic images. Test questions for residents are then systematically designed to ensure an optimal distribution of disease classes and a diverse range of image difficulty levels. In subsequent tests, by analyzing each resident's mistakes in previous tests and finding their error patterns, the evaluation weight of each test image is determined according to each resident's mistakes in previous similar images. During this iterative process, each resident's training period is conducted taking into account their strengths and weaknesses in previous tests. Experimental results demonstrate notable improvements in diagnostic accuracy of residents across various disease types and difficulty levels, along with significant reductions in both false-positive and false-negative rates. The successful performance of this system indicates its usefulness and possibility of being used in personalized education of residents and medical students in other specialties and disease domains, as well.

Keywords: Personalized education, Error prediction, False-negative errors, False-positive errors, Machine learning, Deep learning, Computer vision.

* Corresponding Author's email: hashemzadeh@azaruniv.ac.ir

شخصی‌سازی فرایند آموزش رزیدنت‌های رادیولوژی با استفاده از یادگیری عمیق و استخراج تعاملی مدل خطاهای تشخیصی آنان

سید علی کیان‌مهر^۱، مهدی هاشم‌زاده^{۲،*}

^۱ دانشکده فناوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز، ایران
^۲ آزمایشگاه تحقیقاتی هوش مصنوعی و یادگیری ماشین، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

تاریخ دریافت: ۱۴۰۴/۱۱/۰۱ تاریخ بازبینی: ۱۴۰۴/۰۶/۰۸ تاریخ پذیرش: ۱۴۰۴/۰۷/۳۰

نوع مقاله: پژوهشی

چکیده

تصاویر رادیولوژی از ابزارهای رایج و پرکاربرد تشخیصی هستند که به طور گسترده توسط پزشکان برای تشخیص و نظارت بر روند درمان بیماری‌های مختلف استفاده می‌شوند. تفسیر صحیح این تصاویر نیازمند آموزش رزیدنت‌ها و کارشناسان متخصص با دقت تشخیصی بالا است. در این پژوهش، یک سامانه هوشمند، مبتنی بر یادگیری ماشین، طراحی، پیاده‌سازی و ارزیابی می‌شود که با استفاده از استخراج تعاملی مدل خطاهای تشخیصی مثبت کاذب و منفی کاذب رزیدنت‌های تحت آموزش، دوره آموزش آنها را بصورت شخصی‌سازی شده انجام دهد. راهکار پیشنهادی در قالب یک سامانه برخط آموزش شخصی‌سازی شده طراحی می‌شود که طی فرایندی تکرارشونده، رزیدنت‌ها را در آزمون‌های تطبیقی طراحی شده توسط الگوریتم انتخاب سؤالات، مبتنی بر عملکرد هر رزیدنت، ارزیابی کرده و مدل خطاهای تشخیصی آنها را استخراج می‌کند. در این سامانه، با بهره‌گیری از یک مدل مبتنی بر یادگیری عمیق، ویژگی درجه سختی هر یک از تصاویر رادیوگرافی آموزشی استخراج می‌شود. سپس با رعایت پراکندگی مناسب انواع کلاس‌های بیماری و تنوع درجه سختی تصاویر، سؤالات آزمون‌های رزیدنت‌ها طراحی می‌شود. در آزمون‌های بعدی، با تحلیل خطاهای هر فرد در آزمون‌های قبلی و الگویابی آنها، وزن ارزیابی هر تصویر مطابق با خطاهای تشخیصی پیشین هر فرد در تصاویر مشابه قبلی، مشخص می‌شود. در طی این فرایند تکرارشونده، دوره آموزش هر رزیدنت با در نظر گرفتن نقاط ضعف و قوت او در آزمون‌های قبلی طی می‌شود. نتایج آزمایش‌های انجام شده، نشان‌دهنده بهبود عملکرد کاربران سامانه، هم از نظر دقت تشخیص آنها در هر کلاس با هر درجه سختی تصویر، و هم از نظر کاهش نرخ خطاهای مثبت کاذب و منفی کاذب آنها است. عملکرد موفق و کارآمد این سامانه حاکی از سودمندی و امکان بکارگیری آن در آموزش هوشمند و شخصی‌سازی شده رزیدنت‌ها و دانشجویان پزشکی در دیگر تخصص‌ها و شاخه‌های مختلف انواع بیماری‌ها نیز است.

کلیدواژه‌گان: شخصی‌سازی آموزش، پیش‌بینی خطا، خطاهای منفی کاذب، خطاهای مثبت کاذب، یادگیری ماشین، یادگیری عمیق، بینایی ماشین.

* رایانامه نویسنده مسؤول: hashemzadeh@azaruniv.ac.ir

۱- مقدمه

زبان‌های طبیعی [۶]، ورود و توسعه پیدا کرده است.

در سیستم‌های نوین آموزشی، سامانه‌های کامپیوتری در دو بخش آموزش سنتی و شخصی‌سازی شده توسعه داده شده‌اند. آموزش هدفمند رزیدنت‌های رادیولوژی یکی از عوامل مهم در کاهش خطاهای تشخیصی است. روش‌های سنتی آموزش نمی‌توانند به‌خوبی نیازهای فردی هر رزیدنت را برآورده کنند و به تمام جنبه‌های خطاهای رایج در تفسیر تصاویر پردازند. به‌عبارتی این سیستم‌ها بر اساس مدل‌سازی کاربر کار نمی‌کنند و معمولاً یک الگوی ثابت آموزشی دارند و بدون در نظر گرفتن تفاوت‌های فردی هر رزیدنت، امکانات و محتویات آموزشی یکسانی را برای تمامی رزیدنت‌ها در حوزه‌های مختلف رادیولوژی ارائه می‌دهند [۷].

سامانه‌های شخصی‌سازی آموزش بر اساس این فرضیه طراحی شده‌اند که بسیاری از خطاهای تشخیصی تصادفی نیستند و می‌توان با استخراج الگوهای خطاهای گذشته، آموزش هر فرد را بهبود بخشید. با استفاده از فنون مختلف یادگیری ماشین، می‌توان خطاهای احتمالی آینده را پیش‌بینی و بر اساس آن‌ها برنامه‌های آموزشی شخصی‌سازی‌شده‌ای ارائه کرد که به بهبود نقاط ضعف رزیدنت‌ها کمک کند. با این حال، سامانه‌های آموزش به کمک رایانه مبتنی بر مدل‌سازی کاربر که تاکنون توسعه یافته‌اند، با چالش‌هایی نیز روبرو هستند. این چالش‌ها شامل معیارهای پیچیده داده‌ای، محدودیت در اندازه و تنوع داده‌ها، محدودیت در تعداد و نوع توصیفگرهای ورودی، و نبود یکپارچگی در سامانه برای پیش‌بینی و اصلاح انواع مختلف خطاهای تشخیصی است [۱] و [۶] تا [۱۱].

در این پژوهش، یک سامانه برخط آموزش شخصی‌سازی شده تصاویر پزشکی برای بهبود مهارت‌های رزیدنت‌ها و متخصصان رادیولوژی با هدف کاهش خطاهای تشخیصی مثبت کاذب و منفی کاذب که ترکیبی از یک شبکه عصبی کانولوشنی عمیق و یک الگوریتم آزمون تطبیقی و تعاملی است ارائه می‌شود. سامانه مدنظر با طراحی فرایند محور، رزیدنت‌های رادیولوژی را در یک چرخه‌ی تکرار شونده قرار می‌دهد که با طی کردن مراحل مختلف در قالب آزمون‌های طراحی شده توسط سامانه، مهارت و دقت تشخیصی خود را افزایش دهند. نتایج آزمایشات انجام شده بر روی یک پایگاه تصاویر جامع رادیوگرافی قفسه سینه، حاوی سه کلاس بیماری کوید-۱۹، ذات‌الریه و ریه سالم، و بررسی کارایی سامانه پیشنهادی بر اساس آزمون‌های کاربران واقعی از آن نشان می‌دهد کاربران با تکرار آموزش‌ها و آزمون‌های سامانه و دریافت بازخوردهای تعاملی،

تشخیص به‌موقع و دقیق یک بیماری، مهمترین اقدام در پیشگیری و درمان آن است. خطاها و اختلاف نظرها در تفسیر تصاویر رادیولوژی، مشکلی رایج است که طبق آمار نرخ روزانه آن حدود ۳ تا ۵ درصد تخمین زده شده است. در بررسی هدفمند گزارشات این درصد به مراتب بالاتر نیز است [۱] و [۲]. این خطاها و اختلاف نظرها می‌توانند ناشی از عوامل سیستمی و انسانی باشند، بنابراین نمی‌توان تمام خطاها را به سهل‌انگاری رادیولوژیست‌ها نسبت داد. علاوه بر این، اختلافات در تفسیر تصاویر ممکن است به دلیل تفاوت در دیدگاه‌ها و رویکردهای رادیولوژیست‌ها نیز باشد. برای کاهش این خطاها، لازم است که علل دقیق آن‌ها شناسایی و در هر دو بُعد سیستمی و انسانی اقداماتی اصلاحی انجام شود.

رادیولوژی دارای روش‌های مختلفی همچون رادیوگرافی^۱، فلوروسکوپی^۲ و سونوگرافی^۳ است. تصویربرداری اشعه ایکس یا همان رادیوگرافی، به دلیل در دسترس بودن آن حتی در مناطق کم‌برخوردار و همچنین هزینه‌ی پایین‌تر نسبت به دیگر روش‌های موجود، اغلب در تشخیص اورژانسی و اولیه‌ی اغلب بیماری‌ها، جزء رایج‌ترین ابزارهای تشخیصی بوده و مورد توجه و استفاده پزشکان قرار دارد [۳]. تصاویر اشعه ایکس قفسه سینه برای بررسی و ارزیابی وضعیت ریه‌ها، قلب و دیواره قفسه سینه و همچنین برای کمک به تشخیص علت تنگی نفس، سرفه‌های مزمن و درد قفسه سینه مورد استفاده قرار می‌گیرد. این تصاویر ممکن است برای کمک به تشخیص و نظارت بر روند درمان انواع بیماری‌های ریوی مانند ذات‌الریه، آمفیزم و سرطان استفاده شود. همه‌ی این موارد نیازمند رادیولوژیستی متخصص و نخبه با دقت تشخیصی مناسب برای تشخیص انواع عارضه‌های موجود در تصویر اشعه ایکس است، تا مرتبط‌ترین و مناسب‌ترین نظر کارشناسی ارائه گردد.

امروزه حوزه‌ی رادیولوژی، همچون بسیاری از حوزه‌های کاربردی دیگر، به دنبال ایجاد تحول و گسترش با استفاده هرچه بیشتر از هوش مصنوعی و ابزارهای آن است. به عنوان مثال، در حوزه‌ی تصاویر اشعه ایکس قفسه سینه، مدل‌های یادگیری ماشین در امر تشخیص در سه بُعد^۱ تشخیص مستقل خودکار با استفاده از روش‌های طبقه‌بندی به صورت تشخیص باینری [۴] و چند کلاسه [۳]،^۲ دستیار رادیولوژیست (به صورت ارائه‌ی پیشنهاد کلاسی خاص و یا ناحیه‌ای مشکوک) با استفاده از روش‌های قطعه‌بندی [۵]،^۳ و تفسیر گزارش رادیولوژیست با استفاده از تکنیک‌های پردازش

³ Sonography

¹ Radiography

² Fluoroscopy

آموزشی در مرحله بکارگیری عملی این رزیدنت‌ها بسیار جبران‌ناپذیر می‌تواند باشد.

۲-۲- سیستم‌های آموزش شخصی‌سازی شده

سامانه‌های آموزشی شخصی‌سازی شده با تحلیل دقیق عملکرد هر رزیدنت، آموزش‌هایی متناسب با نقاط قوت و ضعف فردی ارائه می‌دهند. این سامانه‌ها با شناسایی خطاهای تشخیصی و تحلیل الگوهای یادگیری، قادرند برنامه‌های آموزشی ویژه‌ای برای هر کاربر تنظیم کنند. به‌عنوان مثال، اگر رزیدنتی در تشخیص توده‌های خاص دچار مشکل باشد، سیستم قادر به پیش‌بینی این خطاها در تصاویر ورودی جدید خواهد بود. این رویکرد بصورت تدریجی در طی دوره آموزشی باعث بهبود دقت تشخیصی و کاهش خطاهای پزشکی می‌شود. تاکنون برخی پژوهش‌ها در این خصوص در ارتباط با حوزه آموزش رادیولوژیست‌ها انجام شده است که در همه این پژوهش‌ها [۸] تا [۱۰] و [۱۲] تا [۱۸] تصاویر ماموگرافی^۳ به عنوان مورد مطالعه انتخاب شده و فقط در مرجع [۱۴] علاوه بر تصاویر ماموگرافی، تصاویر سی‌تی اسکن^۴ ریه نیز مورد استفاده قرار گرفته است. این پژوهش‌ها به چهار دسته قابل دسته‌بندی هستند: (۱) استفاده از توصیفگرهای BI-RADS^۵ و ویژگی‌های تصویر، (۲) استفاده از ویژگی‌های تصویر با ترکیب ردیابی حرکت چشم، (۳) روش فیلترسازی مشارکتی، و (۴) روش فیلترسازی مشارکتی تقویت‌شده با محتوا^۶. در ادامه به بررسی هر یک از این دسته‌ها و پژوهش‌های انجام شده در هر دسته می‌پردازیم.

۲-۲-۱- استفاده از توصیفگرهای BI-RADS و ویژگی‌های

تصویر

پژوهش انجام شده در مرجع [۸] که توسط مژوروسکی و همکاران در سال ۲۰۱۰ منتشر شد، رویکردی پیشگامانه برای آموزش رایانه‌ای فردی در ماموگرافی از طریق مدل‌سازی کاربران ارائه داد. این مطالعه بر این فرضیه استوار بود که رادیولوژیست‌های در حال آموزش معمولاً الگوهای تکرارشونده‌ای در خطاهای تشخیصی خود دارند که می‌توان این الگوها را شناسایی و پیش‌بینی کرد. آنها با ساخت مدل‌های کاربر برای رزیدنت‌های رادیولوژی بر اساس ویژگی‌های تصویری BI-RADS، به‌ویژه در زمینه تشخیص توده‌های پستان در ماموگرافی، این فرضیه را آزمایش کردند. این مطالعه نشان داد که نتایج به طور آماری معنادار است و مدل‌های پیشنهادی

به تدریج دقت عملکرد خود را در تشخیص بیماری‌ها بهبود داده و میزان خطاهای مثبت کاذب و منفی کاذب خود را کاهش می‌دهند. همچنین، تحلیل دقیق نتایج آزمون‌ها به تفکیک هر کلاس و درجه سختی سؤالات، نشان دهنده تأثیر مثبت این سامانه آموزشی شخصی‌سازی شده بر یادگیری و توانمندسازی رزیدنت‌ها است.

در ادامه، در بخش دوم پژوهش‌های پیشین در حوزه‌ی شخصی‌سازی آموزش، به‌ویژه رزیدنت‌های رادیولوژی، مرور می‌شوند. در بخش سوم جزئیات طراحی و پیاده‌سازی سامانه پیشنهادی ارائه می‌شود. در بخش چهارم نتایج آزمون‌ها و آزمایش‌های انجام‌شده برای ارزیابی کارایی روش پیشنهادی گزارش می‌شود، و نهایتاً در بخش پنجم نتیجه کلی پژوهش ارائه شده و در مورد کارهای آتی بحث می‌شود.

۲- کارهای مرتبط پیشین

با گسترش هوش مصنوعی در حوزه‌های مختلف به‌ویژه در بخش سلامت و تشخیص بیماری‌ها، بر اساس پژوهش‌های منتشر شده در دهه اخیر، تحقیقات در زمینه تشخیص بیماری‌ها از روی تصاویر پزشکی به دو گروه کلی تقسیم می‌شوند: (۱) سیستم‌های تشخیصی و شناسایی مبتنی بر کامپیوتر^۱ (CAD) و (۲) سیستم‌های کمک‌آموزشی مبتنی بر کامپیوتر^۲ (CAE). علاوه بر توسعه سیستم‌های CAD که نقش حیاتی در تشخیص بیماری‌ها دارند، طراحی سیستم‌های CAE برای بهبود عملکرد و مهارت‌های پزشکان و مخصوصاً رادیولوژیست‌ها از اهمیت ویژه‌ای برخوردار است. سیستم‌های CAE از لحاظ استراتژی آموزشی به دو رویکرد سنتی و شخصی‌سازی شده، به شرح ذیل، تقسیم می‌شوند.

۲-۱- سیستم‌های آموزش سنتی

در سیستم‌های آموزشی سنتی یا «یکسان برای همه»، محتوای آموزشی از پیش تعیین شده و به‌صورت یکسان برای همه‌ی کاربران ارائه می‌شود. در چنین رویکردهایی، آموزش بیشتر به تلاش فردی هر کارآموز وابسته است و کمترین توجه به نیازهای خاص هر فرد داده می‌شود. برای مثال، رزیدنتی که در تشخیص یک ناهنجاری خاص ماهرتر است، همان آموزشی را دریافت می‌کند که رزیدنتی با ضعف در آن زمینه دریافت می‌کند، و برعکس. این مسئله باعث می‌شود اغلب رزیدنت‌ها نتوانند در دوره آموزشی خویش به‌طور صحیح و کامل نقاط ضعف خود را برطرف کنند و از زمان دوره آموزشی خود بطور بهینه استفاده نمایند. عواقب چنین سیستم‌های

⁴ CT Scan

⁵ Breast Imaging-Reporting and Data System

⁶ Content-Boosted Collaborative Filtering (CBCF)

¹ Computer Aided Diagnosis (CAD)

² Computer-aided education (CAE)

³ Mammography

داده‌های مورد نیاز توسط سه نفر متخصص خبره برچسب‌گذاری شدند. از رزیدنت‌ها برای مدل‌سازی و ارزیابی سیستم هدف استفاده شده و خطای آنها، که بر اساس محاسبه میزان خطای تشخیص رزیدنت از تشخیص متخصصین بود، محاسبه گردید. در نهایت الگوریتم پیش‌بینی خطا، در سه مرحله‌ی متوالی ناحیه‌بندی تصویر، استخراج ویژگی و مدل‌سازی پیش‌بینی کننده‌ی خطا پیاده‌سازی شد. نتایج نشان داد که الگوریتم ارائه شده به طور موفقیت‌آمیزی موارد دشوار را شناسایی می‌کند و رویکردی امیدوارکننده برای بهبود فرایند آموزشی با تمرکز بر دشوارترین موارد برای هر کارآموز ارائه می‌دهد و از این طریق کارایی آموزش را افزایش می‌دهد.

در پژوهش [۹]، هدف تکمیل پژوهش ارائه شده در مرجع [۱۰] بود. اساس مراحل تا بخش مدل‌سازی خطا در این دو پژوهش یکسان بوده است؛ اما هدف جدید پیش‌بینی خطای مثبت کاذب بود که در پژوهش قبلی انجام نشده بود. الگوریتم پیش‌بینی خطا، در چهار مرحله‌ی متوالی، یعنی خوشه‌بندی برای استخراج مکان‌های کاندید مثبت کاذب، ناحیه‌بندی تصویر، استخراج ویژگی، و ساخت مدل پیش‌بینی کننده پیاده‌سازی شد. نتایج نشان داد الگوریتم پیشنهادی در پیش‌بینی نقاط غیرعادی که کارآموز احتمالاً آن‌ها را از دست خواهد داد، بهبود می‌بخشد. در واقع، این روش به پتانسیل ادغام فن‌های کمک کامپیوتری در محیط‌های آموزشی برای هدف‌گذاری نقاط ضعف خاص کارآموزان و در نتیجه بهبود دقت کلی تشخیصی تمرکز داشت.

۲-۲-۲- استفاده از ویژگی‌های تصویر با ترکیب ردیابی

حرکت چشم

وُیسین و همکاران در سال ۲۰۱۳ [۱۳] به بررسی امکان پیش‌بینی خطاهای تشخیصی در ماموگرافی با ادغام فناوری پیگیری چشم^۱ با ویژگی‌های تصاویر که توسط کامپیوتر و انسان استخراج شده، پرداختند. آنها آزمایش‌هایی با رادیولوژیست‌هایی با سطوح تجربه متفاوت انجام دادند و مدل‌های یادگیری ماشین پیش‌بینی گروهی و شخصی‌سازی شده توسعه دادند. نتایج نشان داد در حالی که مدل‌های گروهی دقت معقولی ارائه دادند، مدل‌های شخصی‌سازی شده، به‌ویژه برای رادیولوژیست‌های باتجربه، موثرتر بودند. گندم‌کار و همکاران در سال ۲۰۱۶ روشی پیشنهاد دادند که این عناصر را برای طبقه‌بندی نواحی مشکوک شناسایی شده در ماموگرافی به‌عنوان مثبت حقیقی (TP) یا مثبت کاذب (FP) ترکیب می‌کرد [۱۸]. با ردیابی حرکات چشم رادیولوژیست‌ها و خوشه‌بندی

می‌توانند تا حدودی سطح دشواری موارد را بر اساس ویژگی‌های تصویری پیش‌بینی کنند که برای کاربردهای عملی حیاتی است. از معایب روش‌های پیاده‌سازی شده در این پژوهش، استفاده از تنها دو ویژگی تصویری و بررسی روش پیشنهادی فقط با ۳۰ تصویر ماموگرافی اشاره کرد. همچنین ویژگی‌های تصویری استفاده شده، توسط کارشناسان بصورت دستی استخراج و ارائه شده بود. این مطالعه پایه‌ای برای توسعه سامانه‌های آموزشی تطبیقی فراهم کرد که می‌توانست به نیازها و چالش‌های خاص هر رادیولوژیست به طور فردی پاسخ دهد.

مژوروسکی و همکاران در سال ۲۰۱۴ در ادامه پژوهش قبلی خویش، این بار مرحله‌ی استخراج ویژگی‌ها از ماموگرافی‌ها را به صورت کاملاً خودکار بجای استفاده از ویژگی‌های انسانی به کار گرفتند [۱۵]. با استفاده از رگرسیون لُجستیک بر روی این ویژگی‌های استخراج شده توسط کامپیوتر، آنان سعی کردند پیش‌بینی کنند که آیا یک رزیدنت به درستی یک توده ماموگرافی را شناسایی خواهد کرد یا خیر. این مطالعه با استفاده از داده‌های سه کارشناس و سه رزیدنت انجام شد و نشان داد که مدل‌های آن‌ها می‌توانند خطاها را با دقتی بهتر از شانس تصادفی پیش‌بینی کنند.

در مطالعه‌ای که توسط ژانگ و همکاران در سال ۲۰۱۴ انجام شد [۱۶]، پژوهشگران به بررسی رویکردی نوین برای بهبود آموزش رادیولوژی از طریق الگوریتم‌های بینایی کامپیوتر پرداختند. این تحقیق بر پیش‌بینی احتمال از دست دادن توده‌های ماموگرافی توسط کارآموزان رادیولوژی با استفاده از ویژگی‌هایی که به طور خودکار از تصاویر استخراج شده بودند، متمرکز بود. این مطالعه بر اساس کارهای پیشین نویسندگان که شامل مدل‌سازی اشتباهات کارآموزان بر اساس ویژگی‌های ارزیابی شده توسط انسان بود، گسترش یافته و از ویژگی‌های استخراج شده توسط کامپیوتر استفاده می‌کرد. با ایجاد مدل‌های فردی برای هر کارآموز، سامانه می‌توانست موارد چالشی که احتمالاً آشکار و ملموس نمی‌شوند را شناسایی کند و بدین ترتیب امکان آموزش کارآمدتر را فراهم می‌کرد. نتایج آزمایشات نشان داد که مدل‌ها توانستند بهتر از شانس تصادفی خطاهای کارآموزان را پیش‌بینی کنند.

تمامی مطالعاتی که تا کنون بحث شد بر روی ماموگرافی معمولی بوده‌اند؛ اما رویکردهای ارائه شده توسط ونگ و همکاران در مراجع [۹] و [۱۰]، روی ماموگرافی دیجیتال انجام شده است. هدف مرجع [۱۰]، استخراج یک مدل کاربر بود که احتمال از دست رفتن و عدم تشخیص یک ضایعه توسط رزیدنت رادیولوژی پیش‌بینی کند. ابتدا

¹ Eye-Tracking

بدون بازخورد کافی^۷ مواجه‌اند که دقت پیش‌بینی را کاهش می‌دهد. برای رفع این چالش‌ها، روش ترکیبی که CBF و CF را با هم ادغام می‌کند، توسط هونگ‌لی لین و همکاران پیشنهاد شد [۱۴]. این روش با بهره‌گیری از نقاط قوت هر دو تکنیک و کاهش نقاط ضعف آن‌ها، بهبودهایی را ارائه داد. الگوریتم فیلترسازی مشارکتی تقویت‌شده با محتوا، دقت پیش‌بینی را با تولید یک ماتریس متراکم ساختگی کارآموز-دسته‌بندی^۸ با استفاده از CBF و سپس اعمال CF برای پیش‌بینی‌های نهایی، افزایش می‌دهد. این رویکرد به طور مؤثری مشکلات کارآموز و موارد ناآشنا را حل می‌کند و آن را به یک راه‌حل امیدبخش برای توسعه سامانه‌های آموزشی شخصی‌سازی شده در آموزش رادیولوژی تبدیل می‌کند.

هریک از مطالعات ذکر شده در هر یک از دسته‌ها دارای موفقیت‌ها و کاستی‌هایی بودند. روش‌های دسته اول غیراحتمالی بودن خطاهای رزیدنت‌ها را اثبات نمود؛ اما کاستی‌هایی داشت. از جمله تعداد کم نمونه و امکان رخ دادن بیش‌برازش، استفاده از تعداد کم توصیفگرهای BI-RADS، تعلق مقادیر گسسته و اسمی به توصیفگرها به‌جای مقادیر پیوسته که منجر به دقت کم، ابهام و پیچیدگی در تشخیص نهایی شده بود. روش‌های دسته دوم، توانست با تلفیق ردیابی چشم رزیدنت‌ها با مقاردهی توصیفگرهای BI-RADS به دقت خوبی در پیش‌بینی و مدل‌سازی خطا برسد و مدل مناسبی به دست آورد. اما عیب‌های مهمی داشت؛ از جمله دشواری پیاده‌سازی در هر محیط آموزشی با هر سطح از امکانات، وابسته بودن به پروتکل‌های خاص، انجام در محیط آزمایشگاهی به‌جای محیط بالینی ماموگرافی و استفاده از تعداد محدودی از توصیفگرها. در مورد سابقه آزمون‌های تطبیقی نیز، که در سامانه پیشنهادی از این رویکرد بهره گرفته شده است، می‌توان گفت این نوع از آزمون‌ها در اوایل قرن بیستم با آلفرد بینه (روان‌شناسی که در سال ۱۹۰۵ اولین شکل از آزمون تطبیقی را به‌عنوان یک آزمون هوش توسعه داد) آغاز شد [۲۰]. در دهه ۱۹۵۰، کامپیوترها شروع به تبدیل آزمون‌های تطبیقی به چیزی کردند که امروزه به‌عنوان CAT شناخته می‌شوند. کارهای پیشگامانه در دهه‌های ۱۹۷۰ و ۱۹۸۰ پایه‌های CAT را تثبیت کرد و بهبودهای بعدی دقت و کارایی آن را افزایش داد [۲۱]. یکی از پیشرفت‌های مهم، ادغام مدل رَش^۹ با تئوری پاسخ به آیتم (IRT) بود [۲۲]. این مهم به CAT اجازه داد

تثبیت‌های دیداری^۱، این مطالعه معیارهای مختلف مرتبط با حرکات چشم، مانند مدت زمان تمرکز، تعداد تثبیت‌ها، و بزرگی حرکت چشمی را استخراج کرده و آن‌ها را با ویژگی‌های مبتنی بر تصویر مانند شدت، شکل و توصیف‌کننده‌های بافت ترکیب کرد. این ترکیب به ساخت یک بردار ویژگی ۴۵ بعدی منجر شد که برای آموزش ماشین‌های بردار پشتیبان^۲ برای هر رادیولوژیست استفاده شد؛ که این امر امکان پیش‌بینی خطاهای شخصی‌سازی‌شده را فراهم می‌کرد. این روش بهبودهای قابل توجهی در امتیازات JAFROC^۳ برای اکثر رادیولوژیست‌های مورد آزمایش نشان داد که نشان‌دهنده دقت بیشتر در تشخیص بین TP و FP است و در نتیجه نرخ بازخوانی مثبت کاذب کاهش می‌یابد [۱۹].

۲-۲-۳- روش فیلترسازی مشارکتی

مژوروسکی و همکارانش در سال ۲۰۱۱ [۱۲] چارچوبی برای یک سامانه‌ی آموزش کامپیوتری تطبیقی ارائه کردند که از مدل‌سازی کاربران برای تطبیق پروتکل‌های آموزشی با نیازهای خاص رادیولوژیست‌ها استفاده می‌کرد. این سامانه از فیلترسازی مشارکتی^۴ (CF) برای پیش‌بینی سطح دشواری موارد ناشناخته برای کارآموزان بر اساس عملکرد گذشته آن‌ها بهره می‌برد. سامانه‌هایی که از فیلترسازی مشارکتی بهره می‌گیرند، مزایای مهمی چون استفاده از بازخورد کاربران مشابه و یادگیری از تاریخچه تعاملات گسترده را ارائه می‌دهند. این روش به‌ویژه در پیش‌بینی دقیق‌تر موارد ناشناخته و شخصی‌سازی محتوا برای کاربران مؤثر است. اما معایب این روش شامل وابستگی به داده‌های موجود است؛ در صورتی که داده‌های کافی از تعاملات کاربران وجود نداشته باشد (مشکل داده‌های کم)، یا کاربران جدیدی به سامانه اضافه شوند، عملکرد مدل کاهش می‌یابد. همچنین، ممکن است این روش در مواجهه با موارد جدیدی که کاربران دیگر هنوز آن را ارزیابی نکرده‌اند (مشکل «مورد ناآشنا»)، ناکارآمد باشد.

۲-۲-۴- روش فیلترسازی مشارکتی تقویت‌شده با محتوا

هر دو روش فیلترسازی مبتنی بر محتوا^۵ (CBF) و فیلترسازی مشارکتی، دارای محدودیت‌های ذاتی هستند. دو روش رایج فیلترسازی مبتنی بر محتوا و فیلترسازی مشارکتی به‌طور جداگانه با مشکلاتی نظیر کارآموزان جدید با تعاملات محدود^۶ و موارد جدید

⁵ Content-Boosted Filtering (CBF)

⁶ Naive Trainee

⁷ Naive Case

⁸ Dense Pseudo Trainee-Ratings Matrix

⁹ Rasch Model

¹ Clustering Fixations

² SVM

³ Jack-knife Alternative Free-response Receiver Operating Characteristics

⁴ Collaborative Filtering (CF)

یادگیری عمیق، و ۳) سامانه برخط آموزش شخصی‌سازی شده. در ادامه، ابتدا مجموعه تصاویر استفاده شده در این پژوهش معرفی می‌شوند و سپس جزئیات مربوط به این سه بخش ارائه می‌گردد.

۳-۱- مجموعه داده استفاده شده در پژوهش

مجموعه داده تصاویر رادیوگرافی قفسه سینه که در مراحل مختلف سیستم استفاده شده است، ترکیبی از سه مجموعه داده زیر است:

۱) **مجموعه داده با عنوان «Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification» [۲۶] و [۲۷]:** این مجموعه داده، متشکل از تصاویر OCT و تصاویر رادیوگرافی قفسه سینه است. در این پژوهش صرفاً از بخش تصاویر رادیوگرافی قفسه سینه این مجموعه داده استفاده می‌شود که در مجموع ۵۸۵۶ تصویر رادیوگرافی قفسه سینه شامل ۴۲۷۳ تصویر دارای ذات‌الریه (۲۷۸۰ مورد باکتریایی، ۱۴۹۳ مورد ویروسی) و ۱۵۸۳ تصویر با وضعیت سالم و عادی وجود دارد.

۲) **مجموعه داده تصاویر رادیوگرافی قفسه سینه کوید-۱۹:** این مجموعه داده ارائه شده توسط کوهن و همکاران [۲۸] در مخزن منبع باز گیت-هاب به اشتراک گذاشته شده است. تصاویر این مجموعه داده از منابع عمومی و همچنین از طریق جمع‌آوری غیرمستقیم از بیمارستان‌ها و پزشکان گردآوری شده است. این پروژه توسط کمیته اخلاق دانشگاه مونترال تایید شده است.

۳) **مجموعه داده تصاویر رادیوگرافی قفسه سینه کوید-۱۹ ارائه شده توسط چانگ و همکاران [۲۹] و [۳۰]:** این مجموعه داده نیز در مخزن منبع باز گیت-هاب توسط چانگ و همکاران آن از شرکت داروین-ای‌آی کانادا^۵ و گروه تحقیقاتی پردازش تصویر و بینایی، دانشگاه واترلو کانادا^۶ منتشر شده است.

پس از انتخاب و ادغام تصاویر موجود در سه مجموعه داده فوق، مجموعه داده نهایی مورد استفاده در این پژوهش، حاوی ۴۲۷۳ تصویر بیماری ذات‌الریه، ۱۵۸۳ تصویر با وضعیت سالم (طبیعی) و ۵۷۶ تصویر بیماری کوید-۱۹ است. شکل ۲ نمونه‌هایی از پایگاه تصاویر گردآوری شده را نشان می‌دهد.

دشواری سؤالات را به طور تطبیقی با سطح مهارت تخمینی فرد تطبیق دهد و به این ترتیب دقت ارزیابی مهارت را بهبود بخشد.

در دهه ۱۹۹۰، ظهور اینترنت CAT را در دسترس تر کرد و منجر به پذیرش آن توسط آزمون‌های بزرگی مانند GRE و TOEFL شد. درحالی‌که این آزمون‌ها از پارادایم کلاسیک CAT فراتر رفته‌اند، اصول آزمون تطبیقی همچنان پایه مدل‌های آزمون کنونی را شکل می‌دهد. روش‌های آماری مختلفی مانند بیزی^۱، حداکثر احتمال^۲، یا روش‌های اطلاعات حداکثری^۳ برای بهینه‌سازی تجربه آزمون استفاده می‌شوند. در واقع، CAT به یک موضوع غالب در گروه‌های سنجش آموزشی تبدیل شده است. برای مثال، در نشست سالانه شورای ملی سنجش آموزشی در سال ۱۹۹۹، یک چهارم مقالات ارائه شده به CAT اختصاص داشت [۲۳]. تحقیقات فعلی CAT موضوعات گسترده‌ای را شامل می‌شود، از جمله توسعه بانک سؤالات، انتخاب سؤالات، تخمین مهارت‌ها و مسائل مختلف مرتبط با امنیت و اعتبار آزمون. این مسائل برای اطمینان از اینکه CAT همچنان یک ارزیابی قابل اعتماد، معتبر و منصفانه باقی بماند، حیاتی هستند. یادگیری ماشین، به‌ویژه از طریق شاخه‌هایی مانند یادگیری عمیق و یادگیری تقویتی، CAT را با تحلیل‌های پیشرفته بر روی مجموعه داده‌های بزرگ، مدل‌سازی رفتارهای دقیق، و تطبیق انعطاف‌پذیر با محیط‌های آزمون مختلف، متحول می‌کند [۲۴]. با اینکه فناوری یادگیری ماشین در CAT هنوز در مراحل اولیه خود قرار دارد، پتانسیل آن مشهود است و یادگیری عمیق، پردازش زبان طبیعی و یادگیری تقویتی در بهبود ویژگی سؤالات، فرایندهای نمره‌دهی و مدل‌سازی تطبیقی نویدبخش موفقیت‌های بزرگ در آینده هستند [۲۵].

۳- روش پیشنهادی

شکل ۱ ساختار کلی سامانه معرفی شده در این پژوهش را نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود، داده مورد استفاده در این سامانه تصاویر رادیوگرافی قفسه سینه است (که می‌تواند با هر داده پزشکی دیگری جایگزین شود) و فرایند طراحی شده دارای سه بخش اصلی است: (۱) پیش پردازش، (۲) مدل

⁴ <https://github.com/ieee8023/covid-chestxray-dataset>

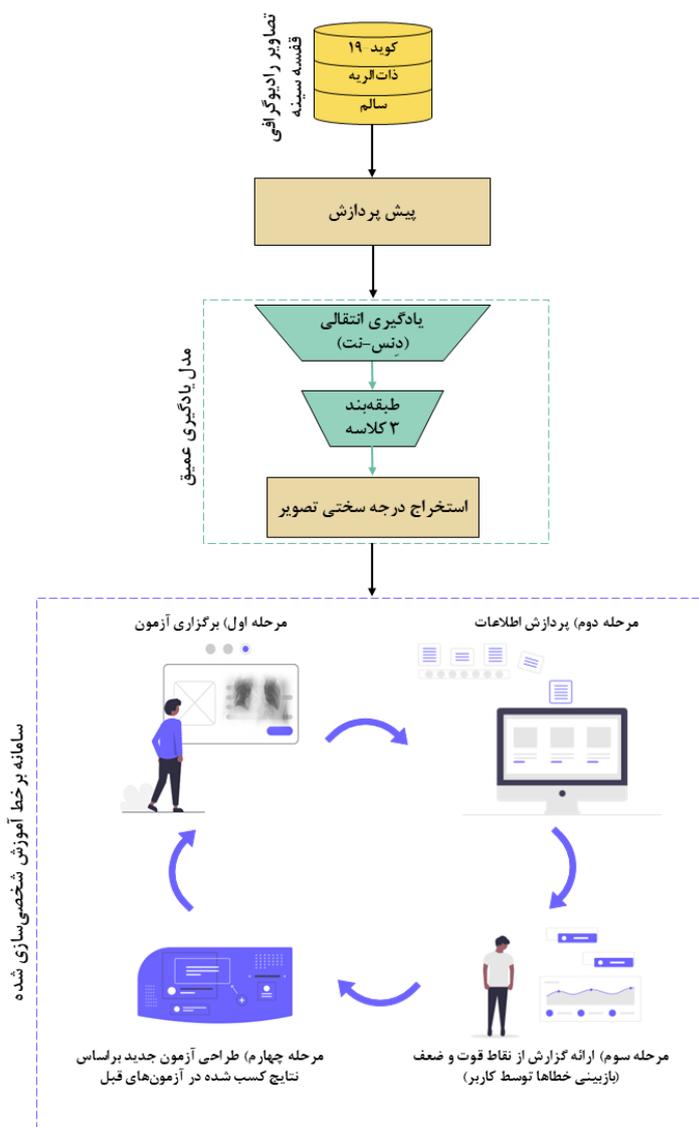
⁵ Canada DarwinAI Corp

⁶ Canada and Vision and Image Processing Research Group, University of Waterloo, Canada

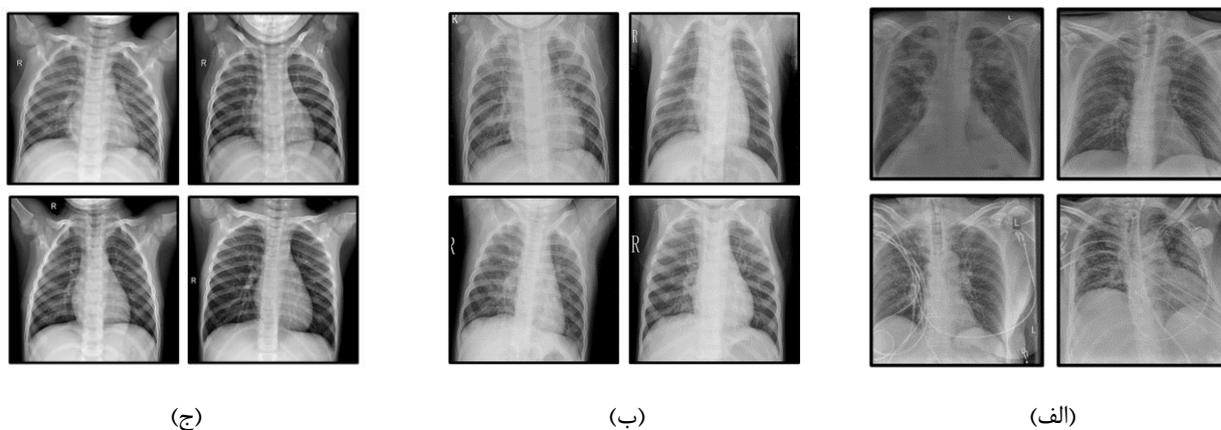
¹ Bayesian

² Maximum likelihood

³ Maximum information methods



شکل ۱. ساختار کلی روش پیشنهادی



شکل ۲. نمونه‌هایی از تصاویر پایگاه‌داده گردآوری شده از کلاس‌های (الف) کویید-۱۹، (ب) ذات‌الریه، و (ج) ریه طبیعی

۲-۳- پیش‌پردازش داده‌ها

با توجه به گردآوری تصاویر از منابع مختلف داده‌ای، ابتدا تصاویر خام جمع‌آوری شده تحت یک سری عملیات آماده‌سازی قرار می‌گیرند تا برای استفاده در مدل یادگیری عمیق مناسب شوند. مرحله پیش‌پردازش شامل عملیات به شرح ذیل است:

۱) **بارگذاری و تغییر اندازه:** تصاویر گردآوری شده به ابعاد ثابت 224×224 پیکسل تغییر اندازه می‌یابند. این اندازه به‌طور خاص با معماری مدل‌های یادگیری انتقالی سازگار است. تغییر اندازه با حفظ نسبت‌ها و کیفیت تصویر انجام می‌شود تا جزئیات مهم برای تحلیل حفظ شود. ۲) **نرمال‌سازی تصاویر:** مقادیر پیکسل‌ها به بازه $[0, 1]$ نرمال می‌شوند. این کار باعث یکنواخت‌سازی داده‌ها، بهبود پایداری عددی و تسریع روند یادگیری مدل‌های عمیق می‌شود.

۳) **برچسب‌گذاری و کدگذاری:** به هر تصویر یک برچسب متنی بر اساس کلاس تعلق می‌گیرد (کوید-۱۹، سالم، ذات‌الریه) که سپس به صورت عددی کدگذاری می‌شود. کلاس ۰ برای کوید-۱۹، کلاس ۱ برای حالت سالم و کلاس ۲ برای تصاویر ذات‌الریه تعریف می‌شود.

۴) **افزایش داده‌ها:** برای بهبود توان تعمیم مدل و کاهش بیش‌برازش، تصاویر آموزشی به صورت تصادفی دستخوش تغییراتی چون چرخش، بزرگ‌نمایی، تغییر کنتراست و انعکاس افقی می‌شوند. این تکنیک‌ها موجب تولید نمونه‌های متنوع و یادگیری ویژگی‌های عمومی‌تر از داده می‌گردد.

۵) **تقسیم داده‌ها:** در نهایت، داده‌ها به سه بخش آموزشی، اعتبارسنجی و ارزیابی تقسیم می‌شوند. این تقسیم‌بندی به‌گونه‌ای انجام می‌شود که نسبت کلاس‌ها در هر سه بخش حفظ شود، تا هر سه بخش نماینده‌ای مناسب از کل مجموعه داده باشد.

۳-۳- مدل یادگیری عمیق

همانطور که در شکل ۱ نمایش داده شد، بخش مدل یادگیری عمیق شامل سه بخش استفاده از تکنیک یادگیری انتقالی از شبکه دِنس-نت، یک طبقه‌بند سه کلاسه، و استخراج درجه سختی تصاویر است. در ادامه هر یک از این بخش‌ها توضیح داده می‌شود.

۱-۳-۳- یادگیری انتقالی دِنس-نت

یادگیری انتقالی یکی از فن‌های مهم در یادگیری عمیق است که در

حوزه بینایی کامپیوتر کاربردهای موفقی داشته است. هدف اصلی یادگیری انتقالی این است که از وزن‌های به دست آمده در یک کاربرد عمومی، که به‌عنوان دانش شناخته می‌شوند، برای بهبود تعمیم‌پذیری در کاربردی دیگر استفاده شود. این روش به ما اجازه می‌دهد تا از مدل‌هایی که بر روی مجموعه داده‌های بزرگ و عمومی آموزش دیده‌اند، برای حل مسائل خاص با داده‌های محدود بهره‌مند شویم [۳۱]. در این پژوهش نیز یادگیری انتقالی به منظور بهره‌برداری از دانش مدل‌های از پیش‌آموزش دیده به کار رفته است. در روش پیشنهادی، از مدل یادگیری انتقالی دِنس-نت [۳۲] استفاده می‌شود. دِنس-نت توسط گائو هوانگ و همکارانش در سال ۲۰۱۷ معرفی شد. این معماری به دلیل نوآوری در اتصال لایه‌ها و کارایی بالا در پردازش تصویر بسیار مورد توجه قرار گرفت. ایده‌ی اصلی دِنس-نت این است که هر لایه در شبکه، به تمام لایه‌های قبلی خود متصل می‌شود. این نوع اتصال منجر به بهبود جریان اطلاعات و گرادیان‌ها در سراسر شبکه می‌شود که به یادگیری بهتر و کارایی بالاتر کمک می‌کند. دِنس-نت-۱۲۱ یکی از نسخه‌های پرکاربرد و متداول دِنس-بلوک است. عدد ۱۲۱ نشان‌دهنده تعداد کل لایه‌های موجود در شبکه است.

۲-۳-۳- طبقه‌بند سه کلاسه

خروجی بخش یادگیری انتقالی به یک طبقه‌بند سه کلاسه، که از یک معماری ساده شبکه عصبی کانولوشنی برخوردار است، داده می‌شود. جزئیات لایه‌های شبکه عصبی عمیق این طبقه‌بند سه طبقه در شکل ۳ نمایش داده شده است. همانطور که در این شکل نشان داده شده است، ابتدا از یک لایه پولینگ میانگین کلی دو بُعدی^۱ استفاده می‌شود. از این لایه به عنوان یک پل بین لایه‌های کانولوشنی و لایه‌های خروجی شبکه دِنس-نت استفاده شده است. در ادامه سه لایه کاملاً متصل^۲ با تعداد نوروهای ۱۲۸، ۶۴، ۱۶ قرار دارد. برای هر یک از این لایه‌ها، از یک لایه حذف تصادفی^۳ با نرخ $0/3$ و تابع فعال‌ساز واحد خطی اصلاح شده^۴ استفاده می‌شود. خروجی آخرین لایه به تابع فعال‌ساز سافت‌مکس^۵ داده می‌شود.

⁴ ReLU

⁵ Softmax

¹ GlobalAveragePooling2D

² Fully Connected Layer

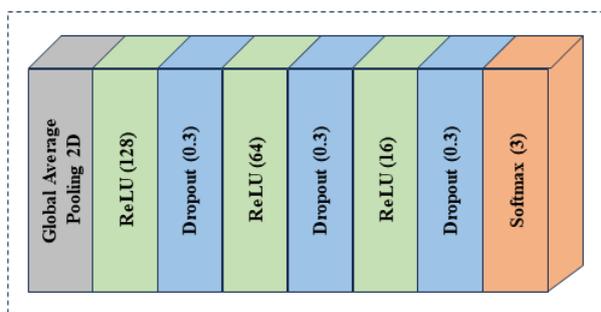
³ Dropout

آزمایش می‌شود. البته بجای استفاده از تشخیص برچسب نهایی مدل، مقادیر احتمالاتی خروجی مدل به‌عنوان میزان اطمینان مدل از پیش‌بینی خود از هر کلاس در نظر گرفته می‌شود. برای محاسبه درجه سختی یک تصویر، احتمال پیش‌بینی شده برای کلاس واقعی آن تصویر (کلاسی که برچسب صحیح تصویر است) در نظر گرفته شده و سپس درجه سختی، که مقدار مکمل این احتمال تعریف می‌شود، محاسبه می‌گردد. به عبارت دیگر درجه سختی هر تصویر $d = 1 - p$ تعریف می‌شود که p برابر است با مقدار احتمال پیش‌بینی شده توسط مدل برای کلاس واقعی هر تصویر. بنابراین، اگر مدل احتمال بالایی برای کلاس واقعی ارائه دهد، مقدار درجه سختی پایین خواهد بود (چون مدل در این پیش‌بینی اطمینان بالایی دارد). برعکس، اگر احتمال برای کلاس واقعی پایین باشد، مقدار درجه سختی بالا خواهد بود (چون مدل در این پیش‌بینی اطمینان کمی دارد).

پس از محاسبه درجه سختی هر یک از تصاویر، مجموعه تصاویر پایگاه به سه گروه آسان (تصاویر با درجه سختی کمتر از ۰/۳۴)، متوسط (تصاویر با درجه سختی بزرگتر یا مساوی ۰/۳۴ و کوچکتر یا مساوی ۰/۶۸)، و سخت (تصاویر با درجه سختی بیشتر از ۰/۶۸) تقسیم می‌شوند (در واقع بازه احتمالاتی ۰ تا ۱ به سه قسمت تقسیم شده است). این تصاویر دسته‌بندی شده از نظر درجه سختی تشخیص در مراحل بعدی و در طراحی ترکیبی سوالات آزمون کاربران خواهند شد.

۳-۴- سامانه برخط آموزش شخصی‌سازی شده

همانطور که در شکل ۱ نشان داده شده است، سامانه برخط آموزش شخصی‌سازی شده دارای چهار مرحله اصلی در فرایند آموزش رزیدنت‌ها است: ۱. مرحله اول: در آزمون اول سوالات به‌صورت متوازن از تمامی کلاس‌ها و از هر نوع درجه سختی به‌صورت تصادفی انتخاب می‌شود. این مرحله با هدف جمع‌آوری داده‌های پایه‌ای برای شناخت اولیه از توانمندی‌ها و ضعف‌های کاربر طراحی شده است؛ ۲. مرحله دوم: بعد از پایان هر آزمون، عملکرد کاربر مشخص شده و اطلاعات آزمون پردازش می‌شود؛ ۳. مرحله سوم: به کاربر گزارشی از نقاط قوت و ضعف او ارائه می‌شود. در این مرحله کاربر امکان بازبینی سوالات را برای بررسی خطاهای خود دارد تا اینکه با نقاط ضعف خود آشنا شود و سوالات آزمون(های) قبل را بررسی کند؛ ۴. مرحله چهارم: آزمون جدید بر اساس نتایج کسب شده توسط کاربر در آزمون(های) قبل با استفاده از الگوریتم آزمون تطبیقی و تعاملی



شکل ۳. جزئیات لایه‌های شبکه عصبی عمیق طراحی شده برای بخش طبقه‌بندی

شبکه عصبی کانولوشنی پیشنهادی با استفاده از داده‌های آموزشی با نرخ یادگیری اولیه ۰/۰۰۱، تعداد ۱۰۰ دور آموزش، تابع هزینه اختلاف آنتروپی موضوعی^۱ و بهینه‌ساز آدام^۲ با نرخ یادگیری ۰/۰۰۱ آموزش داده می‌شود. در مدل پیشنهادی، از بهینه‌ساز آدام استفاده شده که به دلیل کارایی بالا و تنظیمات ساده، به طور گسترده‌ای در شبکه‌های عصبی عمیق استفاده می‌شود.

همچنین، برای جلوگیری از بیش‌برازش، از فن‌های حذف تصادفی با نرخ ۰/۳ و Callbackها استفاده می‌شود. توقف زودهنگام و کاهش نرخ یادگیری، دو نوع از Callbackهایی هستند که در این پژوهش استفاده می‌شوند. تکنیک توقف زودهنگام زمانی که عملکرد مدل در داده‌های اعتبارسنجی بهبود نمی‌یابد، آموزش را زودتر متوقف می‌کند تا از بیش‌برازش جلوگیری شود. تکنیک کاهش نرخ یادگیری نیز در صورتی که بهبودی در تابع هزینه مشاهده نشود، نرخ یادگیری را کاهش می‌دهد تا مدل بتواند با دقت بیشتری به سمت کمینه‌های محلی هدایت شود. کلاس‌های نامتوازن نیز در داده‌های آموزشی با استفاده از وزن‌دهی کلاس‌ها متعادل شده‌اند. در مسائل طبقه‌بندی که داده‌ها نامتوازن هستند، به‌منظور جلوگیری از سوگیری مدل به سمت کلاس‌هایی که تعداد بیشتری از نمونه‌ها را دارند، وزن‌های کلاس‌ها محاسبه می‌شود. این وزن‌ها به مدل کمک می‌کنند تا به هر کلاس اهمیت بیشتری نسبت به میزان فراوانی آن در داده‌های آموزشی بدهد. وزن‌های کلاس در هنگام آموزش به تابع هزینه اضافه می‌شوند تا مدل بر اساس این وزن‌ها آموزش ببیند.

۳-۳-۳- استخراج درجه سختی تصویر

برای استخراج ویژگی درجه سختی هر یک از تصاویر موجود در پایگاه تصاویر استفاده شده در این پژوهش، هر یک از تصاویر پیش‌پردازش شده پایگاه تصاویر به‌عنوان ورودی به مدل آموزش دیده وارد می‌شود. در واقع مدل آموزش دیده با کلیه تصاویر

² Adam

¹ Categorical Cross entropy

درصد از سؤالات از میان تصاویر با درجه سختی متوسط، و ۳۰ درصد باقی‌مانده از میان تصاویر با درجه سختی سخت انتخاب می‌شوند. سؤالات انتخاب شده برای هر کلاس با هم ترکیب می‌شوند و به طور تصادفی مرتب می‌شوند تا ترتیب نهایی سؤالات آزمون مشخص شود. این روش ترکیبی سؤالات آزمون سبب می‌شود که آزمون‌ها از تنوع کافی در سطوح دشواری و کلاس‌ها برخوردار بوده و ترتیب سؤالات به طور تصادفی و بدون الگو مشخص شود که به برقراری عدالت در ارزیابی کاربران کمک می‌کند. رابط‌های کاربری مبتنی بر وب پیاده‌سازی شده برای سامانه پیشنهادی در مرحله برگزاری آزمون در شکل ۴ و شکل ۵ نمایش داده شده است.

۳-۴-۲- مرحله دوم: پردازش اطلاعات

پس از آنکه کاربر به آزمون خود پایان داد، در بخش پردازش اطلاعات بر اساس تشخیص‌های کاربر، معیارهای عملکرد او در آزمون برگزار شده محاسبه می‌شود. در این قسمت ابتدا شمارش تعداد تشخیص‌های مثبت^۱ کاذب^۲، منفی^۳ کاذب^۴ و منفی واقعی^۵ برای هر کلاس انجام شده و ذخیره می‌شود. سپس با استفاده از این اطلاعات، محاسبه‌ی معیارهای ارزیابی صحت^۶، بازیابی^۷، دقت^۸ و امتیاز^۹ برای هر کلاس بصورت مجزا و تجمیعی انجام می‌شود. در ادامه کارنامه کاربر بر اساس این اطلاعات به‌روزرسانی می‌شود. به عبارتی در این مرحله علاوه بر محاسبه‌ی دقیق معیارهای عملکرد برای آزمون، کارنامه کاربر که اطلاعات مهمی همچون وضعیت کاربر و نقاط ضعف و قدرت وی در آن ذخیره شده، به‌روزرسانی می‌شود.

که در ادامه شرح داده خواهد شد، طراحی می‌شود. سؤالات آزمون‌های بعدی، با هدف تمرکز بیشتر بر نقاط ضعف هر کاربر انتخاب می‌شوند. این مراحل تا زمانی که میزان دقت و میزان خطای مثبت^۱ کاذب و منفی^۲ کاذب کاربر در هر سه کلاس مدنظر به مقدار قابل قبولی برسد، تکرار شده و ادامه خواهد داشت. البته در هر تکرار، تمام داده‌های آزمون‌های قبلی برای جلوگیری از سوگیری و تأثیر عوامل حاشیه‌ای لحاظ می‌گردد.

در سامانه‌ی طراحی شده، شخصی‌سازی بر مبنای ترکیبی از تحلیل عملکرد گذشته‌ی هر رزیدنت، الگوی خطاهای او، و به‌کارگیری آزمون تطبیقی و تعاملی انجام می‌گیرد. شخصی‌سازی مدنظر در این سامانه نه فقط انتخاب سؤال ساده یا دشوار، بلکه یک فرایند عمیق تشخیص الگوهای رفتاری و شناختی در تصمیم‌گیری‌های تشخیصی کاربران است. این سیستم با بهره‌گیری از بازخورد مستمر، سعی در بهینه‌سازی فرآیند یادگیری انفرادی و هدایت کاربر به سمت رفع نواقص شناختی خاص خود دارد. در ادامه هر چهار مرحله مذکور به‌صورت کامل تشریح می‌شود.

۳-۴-۱- مرحله اول: برگزاری آزمون

در آزمون اولیه، تعداد کل سؤالات تعیین شده برای آزمون، به طور مساوی بین سه کلاس کوید-۱۹، ذات‌الریه و ریه سالم تقسیم می‌شود. برای هر کلاس، سؤالات بر اساس سطح دشواری به سه گروه آسان، متوسط، و سخت تقسیم می‌شوند. ۳۰ درصد از سؤالات انتخابی برای هر کلاس از میان تصاویر با درجه سختی آسان، ۴۰

⁵ Accuracy

⁶ Recall

⁷ Precision

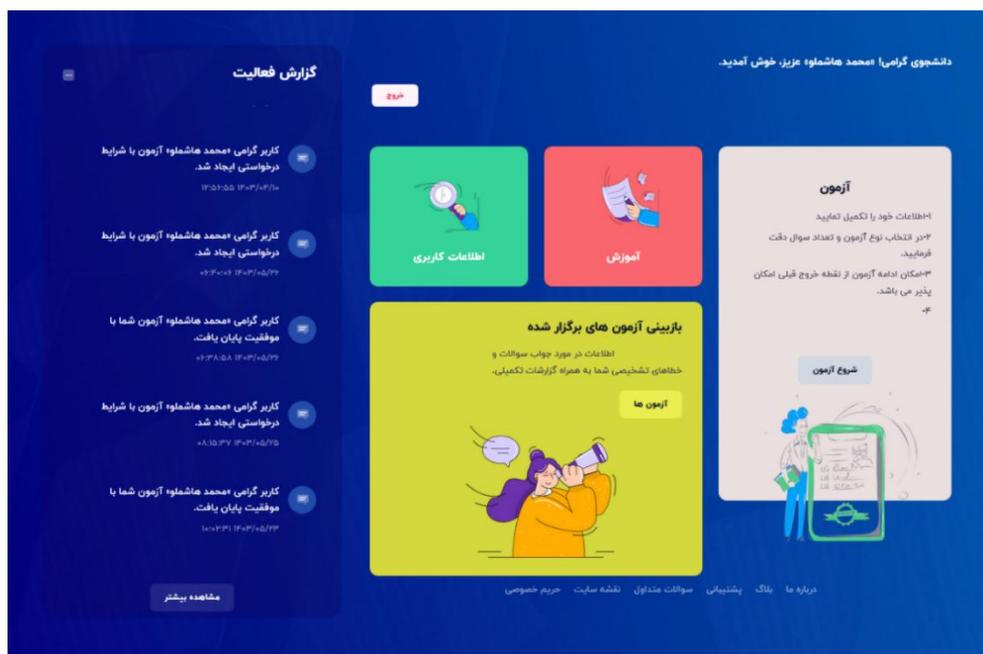
⁸ F-score

¹ False Positive (FP)

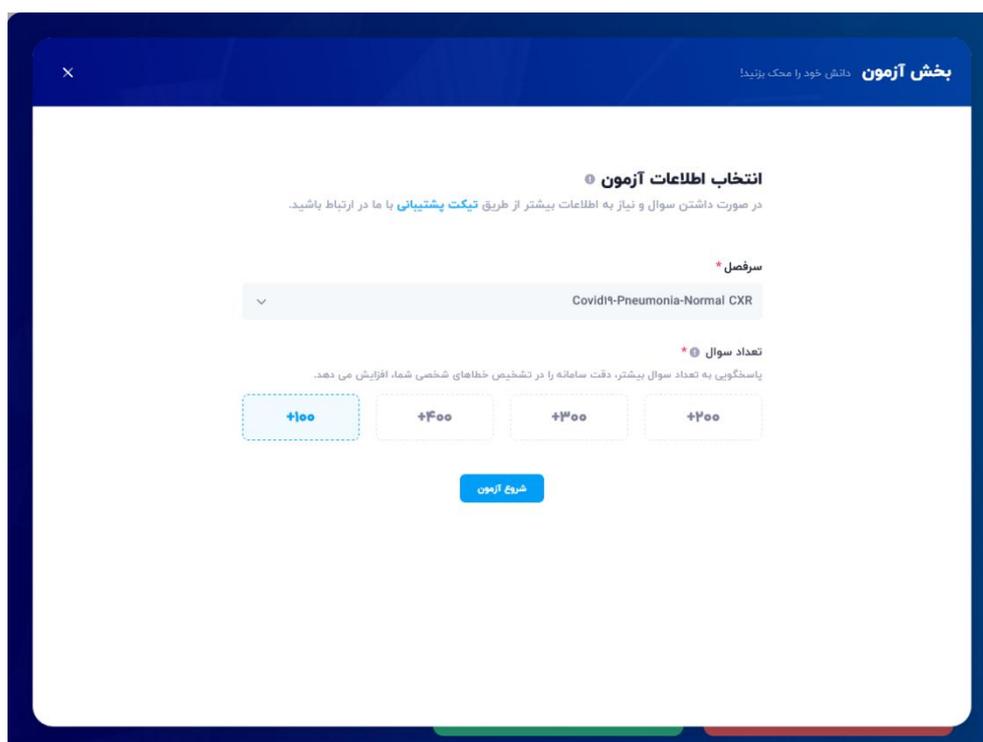
² False Negative (FN)

³ True Positive (TP)

⁴ True Negative (TN)



(الف)



(ب)

شکل ۴. رابط‌های کاربری مبتنی بر وب پیاده‌سازی شده برای مرحله اول سامانه برخط آموزشی. (الف) صفحه پیشخوان هر کاربر، و (ب) بخش انتخاب سر فصل و تعداد سؤالات برای کاربران

شکل ۵. رابط کاربری مبتنی بر وب پیاده‌سازی شده برای صفحه آزمون

ترتیب ماتریس درهم‌ریختگی، معیارها برای هر کلاس، جواب‌های صحیح و جواب‌های غلط را نشان می‌دهد.

از جدول درهم‌ریختگی، کاربر می‌تواند آگاه شود که چه تعداد از پاسخ‌های اشتباهی که برای یک کلاس داده است، مربوط به کدامین کلاس‌ها است. جدول جواب‌های صحیح، اطلاعاتی در مورد اینکه کاربر چه تعداد سؤال از درجات سختی مختلف در کلاس‌های مختلف توانسته پاسخ صحیح دهد، را ارائه می‌دهد. در مورد جدول جواب‌های غلط نیز همان اطلاعات برای سؤالاتی است که کاربر نتوانسته درست پاسخ دهد. این جداول کمک می‌کند تا کاربر یک ارزیابی جامع از میزان دقت و سهل‌انگاری خود به دست آورد.

۳-۴-۳- مرحله سوم: ارائه گزارش از نقاط قوت و ضعف به کاربر

در این بخش، گزارشی اختصاصی برای هر آزمون ارائه می‌شود. هنگامی که کاربر برای بررسی آزمون خود وارد صفحه‌ی گزارش آزمون می‌شود، یک رابط کاربری همانند نمونه موجود در شکل ۶ را مشاهده می‌کند. در این قسمت، گزارش تعداد پاسخ صحیح و غلط کاربر از تعداد کل سؤالات نشان داده شده و به تفکیک هر کلاس دقت کاربر گزارش می‌شود. همچنین برای هر کلاس مشخص می‌شود که از سؤالات انتخاب شده برای هر کلاس، کاربر توانسته چه درصدی از هر نوع درجه سختی را درست پاسخ دهد. در قسمت پائین این رابط کاربری (شکل ۶)، چهار جدول طراحی شده که به

خلاصه عملکرد:
تعداد پاسخ صحیح شما: ۵۳ - تعداد پاسخ غلط شما: ۴۷
تعداد سوال نرمال: ۳۳ تعداد سوال ذات الریه: ۳۳ تعداد سوال کوید-۱۹: ۳۴

تشخیص تصاویر قفسه سینه سالم:
دقت عملکرد شما در تشخیص صحیح تصاویر قفسه سینه سالم ۷۸.۷۹٪ می باشد که دقت متوسطی می باشد.
از سوالات انتخاب شده برای شما ۸۹.۸۸٪ از سوالات آسان، ۵۷.۵۷٪ از سوالات متوسط، ۷۰٪ از سوالات سخت را درست پاسخ داده اید.

تشخیص تصاویر قفسه سینه دارای ذات الریه:
دقت عملکرد شما در تشخیص صحیح تصاویر قفسه سینه دارای ذات الریه ۴۲.۴۲٪ می باشد که دقت ضعیفی تلقی می شود.
از سوالات انتخاب شده برای شما ۵۵.۵۵٪ از سوالات آسان، ۸۵.۸۵٪ از سوالات متوسط، ۱۱.۱۱٪ از سوالات سخت را درست پاسخ داده اید.

تشخیص تصاویر قفسه سینه دارای بیماری کوید-۱۹:
دقت عملکرد شما در تشخیص صحیح تصاویر قفسه سینه دارای بیماری کوید-۱۹ ۳۸.۲۴٪ می باشد که دقت ضعیفی می باشد.
از سوالات انتخاب شده برای شما ۶۷.۶۷٪ از سوالات آسان، ۵۴.۵۴٪ از سوالات متوسط، ۳۳.۳۳٪ از سوالات سخت را درست پاسخ داده اید.

تعداد سوالات: ۱۰۰

- سوال ۱
- سوال ۲
- سوال ۳
- سوال ۴
- سوال ۵
- سوال ۶
- سوال ۷
- سوال ۸
- سوال ۹
- سوال ۱۰
- سوال ۱۱
- سوال ۱۲
- سوال ۱۳
- سوال ۱۴
- سوال ۱۵
- سوال ۱۶
- سوال ۱۷
- سوال ۱۸
- سوال ۱۹
- سوال ۲۰
- سوال ۲۱

ماتریس درهم ریختگی

کلاس / معیار	مثبت واقعی	مثبت کاذب	منفی واقعی	منفی کاذب
سالم	۲۶	۲۵	۴۲	۷
ذات الریه	۱۴	۱۱	۵۶	۱۹
کوید-۱۹	۱۳	۱۱	۵۵	۲۱

معیارها برای هر کلاس

کلاس / معیار	مثبت واقعی	مثبت کاذب	منفی واقعی	منفی کاذب
سالم	۲۶	۲۵	۴۲	۷
ذات الریه	۱۴	۱۱	۵۶	۱۹
کوید-۱۹	۱۳	۱۱	۵۵	۲۱

جواب های صحیح

کلاس	آسان	متوسط	سخت
NORMAL	۸	۱۱	۷
PNEUMONIA	۶	۷	۱
COVID19	۲	۸	۳

جواب های غلط

کلاس	آسان	متوسط	سخت
NORMAL	۱	۳	۳
PNEUMONIA	۵	۶	۸
COVID19	۱۰	۵	۶

شکل ۶. نمونه رابط کاربری پیاده‌سازی شده برای گزارش خلاصه عملکرد هر آزمون در مرحله سوم سامانه برخط آموزشی

هر مرحله از مراحل سامانه که کاربر نیاز به مشاهده نمونه‌هایی از یک کلاس و یا مرور نحوه تشخیص و تمیز سه کلاس از یکدیگر داشته باشد، بتواند به‌صورت خلاصه به اطلاعات آموزشی کافی دسترسی داشته باشد. در شکل ۸ نمونه‌ای از نحوه نمایش اطلاعات در صفحه آموزش نشان داده شده است.

همچنین امکان بررسی سؤال به سؤال برای کاربر فراهم شده است. در شکل ۷ نحوه نمایش هر سؤال برای کاربر نشان‌داده شده است. مشاهده می‌شود که برای کاربر، علاوه بر نمایش صحیح یا غلط بودن پاسخ ارسالی، جواب صحیح و درجه سختی سؤال نیز برای راهنمایی بهتر و توجه به جزئیات مطلوب نمایش داده می‌شود. علاوه بر این اطلاعات، بخشی با عنوان آموزش برای کاربر تدارک دیده شده تا در



سوال ۴: نادرست **درجه سختی: متوسط**

جواب شما: Pneumonia **جواب صحیح: Covid19**



تعداد سوالات: ۱۰۰

- سوال ۱ ✓
- سوال ۲ ✓
- سوال ۳ ✓
- سوال ۴ ✓
- سوال ۵ ✓
- سوال ۶ ✓
- سوال ۷ ✓
- سوال ۸ ✓
- سوال ۹ ✓
- سوال ۱۰ ✓
- سوال ۱۱ ✓
- سوال ۱۲ ✓
- سوال ۱۳ ✓
- سوال ۱۴ ✓
- سوال ۱۵ ✓
- سوال ۱۶ ✓
- سوال ۱۷ ✓
- سوال ۱۸ ✓

دانشجو

Copyright © ۲۰۲۴

درباره ما بلاگ پشتیبانی سوالات متداول نقشه سایت حریم خصوصی

شکل ۷. نمونه رابط کاربری پیاده‌سازی شده برای نمایش سؤالات در قسمت بازبینی مرحله سوم سامانه برخط آموزشی

آموزش تشخیص بیماری های ذات الریه و کووید-۱۹ از تصاویر رادیوگرافی قفسه سینه:

تصویربرداری با اشعه ایکس از قفسه سینه (Chest X-Ray یا CXR) یکی از ابزارهای اصلی برای تشخیص بیماری های ریوی مانند پنومونی و کووید-۱۹ است. در ادامه، جزئیاتی از نحوه تشخیص این بیماری ها بر اساس تصاویر CXR با دقت بیشتری بررسی خواهد شد.

- سایه بندی (Opacity)

ویژگی ها: نواحی سایه بندی به صورت مناطق خاکستری یا سفید در تصاویر CXR دیده می شوند که منعکس کننده تجمع مایعات، سلول های التهابی یا مواد دیگر در آئولولها هستند. **مفهوم:** این سایه بندی نمایانگر التهاب یا عفونت در ریه است. **در پنومونی:** سایه بندی ممکن است به صورت مناطق کوچک یا بزرگ، معمولاً با لبه های نامنظم، دیده شود. این مناطق ممکن است در لبه های خاصی متمرکز شوند. **در کووید-۱۹:** سایه بندی های پراکنده، به ویژه در مناطق زیر پلورال و محیطی (پریفرال) ریه ها، دیده می شود. این الگو ممکن است در هر دو سمت ریه ها ظاهر شود و در مراحل پیشرفته تر بیماری، این سایه بندی ها افزایش یابد.

- کنسولیویداسیون (Consolidation)

ویژگی ها: مناطق کنسولیویداسیون به صورت نواحی سفید و چگالی بالا دیده می شوند. **مفهوم:** این نواحی نشان دهنده تجمع مایع، پروتئین و سلول های التهابی در آئولولها و به دلیل عفونت یا التهاب شدید هستند. **در پنومونی:** کنسولیویداسیون معمولاً به صورت مناطق موضعی و متمرکز دیده می شود که می تواند لوب یا بخش خاصی از ریه را درگیر کند. این الگو به عنوان نشانه ای از عفونت باکتریایی شایع است. **در کووید-۱۹:** کنسولیویداسیون نیز ممکن است دیده شود، اما اغلب به صورت پراکنده تر و در مناطق وسیعتری قرار دارد. این نواحی معمولاً در مراحل پیشرفته عفونت ویروسی افزایش می یابند.

- ضایعات شیشه مات (Ground-Glass Opacities یا GGO)

ویژگی ها: این ضایعات به صورت نواحی نیمه شفاف خاکستری دیده می شوند. **مفهوم:** ضایعات شیشه مات نشان دهنده التهاب بینابینی و التهابات غیرموضعی در آئولولها هستند. **در پنومونی:** این ضایعات ممکن است دیده شوند، اما معمولاً کمتر شایعند و بسته به نوع پنومونی، ممکن است شدت آنها متفاوت باشد. **در کووید-۱۹:** ضایعات شیشه مات یکی از ویژگی های کلیدی کووید-۱۹ در سی تی اسکن و CXR است. این ضایعات اغلب در هر دو لوب ریه و در مناطق زیر پلورال و محیطی دیده می شوند.

- ضخیم شدن پلورا (Pleural Thickening)

ویژگی ها: ضخیم شدن پلورا به صورت نوارهای سفید و متراکم در امتداد جداره ریه ها دیده می شود. **مفهوم:** این حالت ممکن است نشان دهنده التهاب یا عفونت مزمن در پلورا باشد. **در پنومونی:** ممکن است ضخیم شدن پلورا دیده شود، اما معمولاً شایع نیست. **در کووید-۱۹:** ضخیم شدن پلورا ممکن است کمتر دیده شود و این ویژگی به عنوان مشخصه اصلی کووید-۱۹ محسوب نمی شود.

- نکات مهم در تشخیص تفاوت پنومونی و کووید-۱۹ در CXR

الگوی توزیع: در پنومونی باکتریایی، کنسولیویداسیون ها معمولاً به صورت متمرکز و در یک لوب خاص ظاهر می شوند، در حالی که در کووید-۱۹ الگوی ضایعات پراکنده تر و در هر دو لوب دیده می شود. **شدت و پراکندگی ضایعات:** شدت و گستردگی ضایعات شیشه مات و کنسولیویداسیون در کووید-۱۹ به مراتب بیشتر و پراکنده تر است. **مقایسه با سایر روش های تشخیصی:** برای تشخیص دقیق تر، نتایج CXR معمولاً با نتایج سی تی اسکن و شواهد بالینی دیگر مانند آزمایش های PCR مقایسه می شود.

- نتیجه گیری

تصویربرداری با اشعه ایکس از قفسه سینه یکی از روش های مؤثر برای تشخیص بیماری های ریوی از جمله پنومونی و کووید-۱۹ است. با تفسیر دقیق و استفاده از دانش تخصصی برای تشخیص الگوهای مشخصه، پزشکان می توانند این بیماری ها را به موقع تشخیص داده و درمان مناسبی را برای بیماران تجویز کنند.

نمونه تصاویر مربوط به هر بیماری و قفسه سینه سالم

- نمونه های زیر برای یادگیری وضعیت نرمال قفسه سینه در تصویر رادیوگرافی گردآوری شده است:



- نمونه های زیر برای یادگیری وضعیت ذات الریه قفسه سینه در تصویر رادیوگرافی گردآوری شده است:



- نمونه های زیر برای یادگیری وضعیت کووید-۱۹ قفسه سینه در تصویر رادیوگرافی گردآوری شده است:



شکل ۸. نمونه رابط کاربری پیاده سازی شده برای صفحه آموزشی سامانه برخط آموزشی

معیار را معرفی می‌کند.

$$FNR_{class} = \frac{FN_{class}}{TP_{class} + FN_{class}} \quad (1)$$

معیار نرخ مثبت کاذب نیز نشان می‌دهد که چه نسبتی از نمونه‌های منفی واقعی به اشتباه به عنوان مثبت پیش‌بینی شده‌اند. رابطه (۲)، تعریف ریاضی این معیار را نشان می‌دهد.

$$FPR_{class} = \frac{FP_{class}}{FP_{class} + TN_{class}} \quad (2)$$

نهایتاً معیار خطای کاربر در این پژوهش، مجموع دو نرخ منفی کاذب و مثبت کاذب تعریف می‌شود که رابطه (۳) مبین آن است.

$$Error_{class} = FPR_{class} + FNR_{class} \quad (3)$$

مجموع این دو نرخ به دلیل پوشش جامع‌تر خطاهای سامانه، تعادل بین خطاهای مثبت و منفی کاذب، و قابلیت تنظیم و بهینه‌سازی ساده‌تر، یک معیار قدرتمند و مناسب برای ارزیابی عملکرد کاربر در سامانه آموزش شخصی‌سازی شده براساس الگوریتم انتخاب سوال تطبیقی و تعاملی است. برای مثال، در مسئله مورد بحث در این پژوهش، اینکه یک رزیدنت به یک بیمار که دارای بیماری کوید-۱۹ است، تشخیص سالم بدهد (یعنی بیمار دارای کوید-۱۹ نیست که خطای منفی کاذب تلقی می‌شود)، و یا به بیماری که سالم است تشخیص کوید-۱۹ بدهد (خطای مثبت کاذب)، هر دو به یک اندازه خطایی فاحش و گاه جبران‌ناپذیر هستند که می‌تواند پیامدهای جدی به همراه داشته باشد. ذکر این نکته حائز اهمیت است که این معیار برای هر کلاس به صورت جداگانه محاسبه می‌شود.

(۲) تخصیص وزن به کلاس‌ها: بر اساس معیار خطای کاربر، محاسبه شده برای هر کلاس، وزن‌های مربوط به هر کلاس تعیین می‌شوند. این وزن‌ها با استفاده از ترکیب نرخ مثبت کاذب و نرخ منفی کاذب که به عنوان خطای کاربر در نظر گرفته شده است، محاسبه می‌شوند. رابطه (۴) نحوه محاسبه وزن هر کلاس را در انتخاب سؤالات تعریف می‌کند. بر اساس این رابطه، هرچه مقدار خطای یک کلاس، بالاتر باشد، وزن بیشتری به آن در انتخاب سؤالات اختصاص داده می‌شود، به این معنا که سامانه به این کلاس توجه بیشتری خواهد داشت.

۳-۴-۴- مرحله چهارم: طراحی آزمون بر اساس نتایج

کسب شده در آزمون‌های قبل

هدف اصلی سامانه‌های آزمون تطبیقی، بهینه‌سازی فرایند ارزیابی با انتخاب سؤالاتی است که با سطح دانش و توانایی‌های فرد آزمون دهنده هماهنگ باشند. در این پژوهش یک الگوریتم برای انتخاب سؤالات بر اساس عملکرد کاربر در آزمون‌های قبلی ارائه می‌شود، که از فن‌های آماری و تئوری‌های مطرح در آزمون‌سازی استفاده می‌کند. این فنون در مراجع [۳۳] و [۳۴] توضیح داده شده است. آزمون انطباقی کامپیوتری^۱ که از روان‌سنجی سرچشمه می‌گیرد، امکان ارزیابی کارآمد را فراهم می‌کند [۳۵] و [۳۶]. این روش به طور گسترده در سناریوهای برگزاری آزمون از جمله در دولینگو^۲، جی‌آرای^۳، تافل^۴ و جی‌مت^۵ استفاده می‌شود. الگوریتم پیشنهادی برای انتخاب سؤالات بر اساس عملکرد قبلی هر کاربر از ترکیبی از معیارهای عملکرد کاربر و مفاهیم نظریه پاسخ به آیت^۶ و یادگیری تطبیقی^۷ که بخشی از یادگیری تعاملی^۸ است، برای انتخاب بهینه سؤالات استفاده می‌کند. همانطور که در شکل ۹ نشان داده شده است، الگوریتم پیشنهادی از چهار مرحله تشکیل می‌شود که عبارتند از: (۱) محاسبه‌ی معیار خطا، (۲) تخصیص وزن به کلاس‌ها، (۳) انتخاب سؤالات، و (۴) تنظیم نهایی و ارائه‌ی سؤالات. در ادامه، هر یک از این مراحل تشریح می‌شوند.

(۱) محاسبه‌ی معیار خطا: بر اساس اطلاعاتی که در مرحله دوم پردازش و استخراج شده، دو معیار نرخ منفی کاذب^۹ و نرخ مثبت کاذب^{۱۰} محاسبه می‌شود. معیار نرخ منفی کاذب نشان می‌دهد که چه نسبتی از نمونه‌های مثبت واقعی به اشتباه به عنوان منفی پیش‌بینی شده‌اند. در زمینه‌هایی مانند پزشکی، امنیت و یا مسائل حیاتی دیگر، نرخ منفی کاذب باید تا حد ممکن پایین باشد، زیرا از دست دادن موارد مثبت (به‌خصوص در موارد حیاتی مثل بیماری) می‌تواند منجر به پیامدهای جدی شود. به‌عنوان مثال، در یک سامانه تشخیص کوید-۱۹، اگر نرخ منفی کاذب بالا باشد، این بدان معناست که بیماران مبتلا به کوید-۱۹ به اشتباه سالم تشخیص داده شده‌اند و این امر می‌تواند منجر به عدم درمان به‌موقع شود. به همین دلیل، در چنین کاربردهایی، مدل‌ها یا افرادی که نرخ منفی کاذب کمتری دارند، ترجیح داده می‌شوند. رابطه (۱)، تعریف ریاضی این

⁶ Item Response Theory (IRT)

⁷ Adaptive Learning

⁸ Interactive Learning

⁹ False Positive Rate (FPR)

¹⁰ False Negative Rate (FNR)

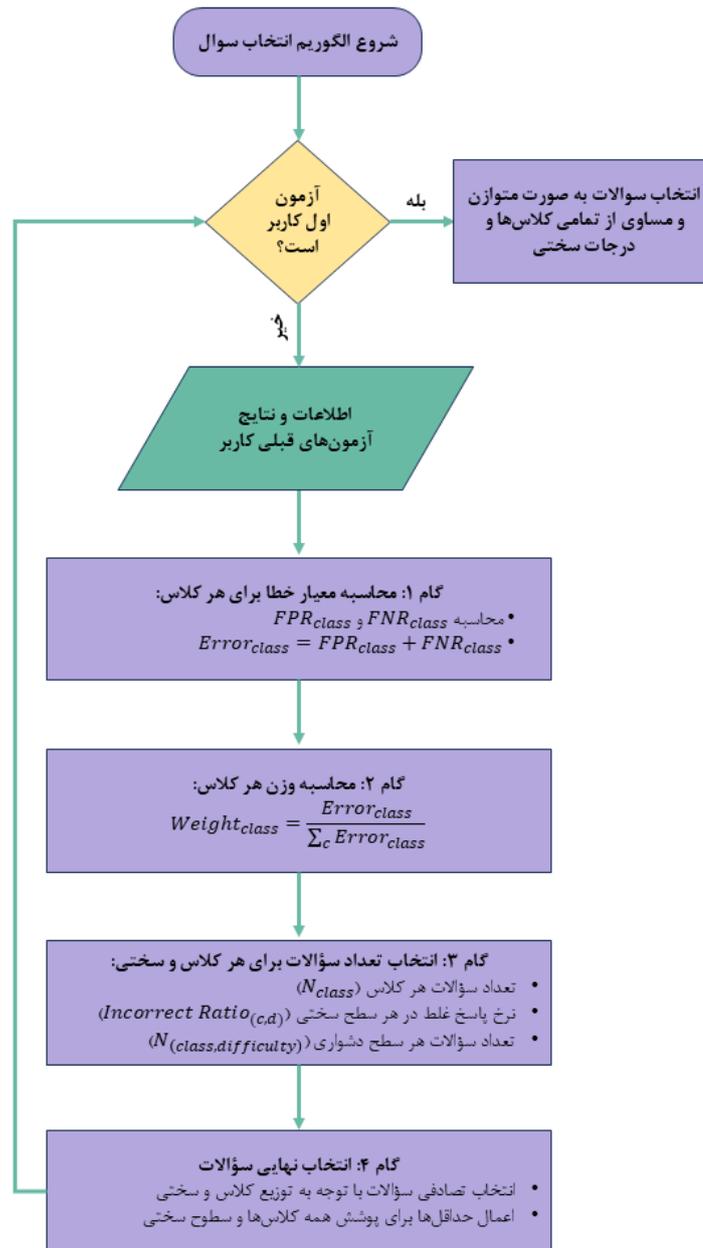
¹ Computerized Adaptive Testing (CAT)

² Duolingo

³ GRE

⁴ TOEFL

⁵ GMAT



شکل ۹. نمای کلی الگوریتم انتخاب سؤال در سامانه برخط آموزش شخصی‌سازی شده

N_{Total} بیانگر تعداد کل سوالات تعیین شده برای آزمون است.

$$N_{class} = \max((20 \times N_{Total})/100, Weight_{class} \times N_{Total}) \quad (5)$$

اگر کاربر بتواند در یک کلاس به دقت بالایی برسد (کاهش قابل توجه میزان خطا)، در حالت عادی این مهم باعث خواهد شد تا وزن آن کلاس کاهش یافته و به تبع آن تعداد سوالات انتخابی از آن کلاس به حداقل برسد. به دلیل آن که آموزش در حوزه‌ی تصاویر پزشکی، تک مرحله‌ای اتفاق نمی‌افتد، این روش ممکن است در چندین مرحله باعث شود کاربر دقت به دست آمده در آن کلاس را

$$Weight_{class} = \frac{Error_{class}}{\sum_c Error_{class}} \quad (4)$$

۳) انتخاب سوالات: تعداد سوالات مورد نیاز از هر کلاس و سطح

دشواری بر اساس وزن‌های محاسبه‌شده و نسبت خطاهای کاربر تعیین می‌شود. الگوریتم به‌گونه‌ای طراحی شده که سوالات را به نحوی انتخاب کند که توازن بین سطح دشواری و عملکرد کاربر در هر کلاس نیز حفظ شود. رابطه (۵) نحوه محاسبه تعداد سوالات انتخابی از هر کلاس بر اساس وزن‌های تعیین شده را تعریف می‌کند. در این رابطه N_{class} نشانگر تعداد سوالات برای هر کلاس بوده و

می‌شود.

لازم به ذکر است که در محاسبات انجام شده، در هر مرحله اطلاعات تمامی آزمون‌های برگزار شده دخیل هستند و معیار برای طراحی آزمون جدید، فقط آخرین آزمون برگزار شده نیست. دلیل این کار جلوگیری از تأثیرگذاری عوامل حاشیه‌ای همچون حواس‌پرتی و خستگی در یک آزمون، بر کل روند آموزشی است.

۴) تنظیم نهایی و ارائه سؤالات: پس از تعیین ترکیب تعداد سؤالات بر اساس وزن‌ها و سطوح دشواری مختلف، سؤالات نهایی به صورت تصادفی از بین سؤالات دسته‌بندی شده در مرحله دوم این الگوریتم انتخاب و به کاربر ارائه می‌شوند. این تصادفی‌سازی در انتخاب سؤالات نهایی کمک می‌کند تا از یکنواختی جلوگیری شود و تنوع در سؤالات حفظ گردد. همچنین، این مرحله به جلوگیری از پیش‌بینی‌پذیری در آزمون کمک می‌کند و امنیت آزمون را افزایش می‌دهد.

۴- آزمایش‌ها و نتایج عملی

در این بخش، کارایی روش پیشنهادی توسط آزمایش‌های عملی مورد ارزیابی قرار می‌گیرد. ابتدا نتایج آزمایش‌های انجام گرفته برای ارزیابی کارایی و دقت مدل تعیین کننده‌ی درجه سختی گزارش می‌شود. سپس نتایج آزمون‌های برگزار شده توسط کاربران واقعی سامانه، تحلیل و بررسی شده و به ارزیابی کارآمدی سامانه معرفی شده پرداخته می‌شود.

۴-۱- ارزیابی مدل یادگیری انتقالی عمیق

در این بخش، به بررسی و تحلیل عملکرد مدل یادگیری انتقالی عمیق در روش پیشنهادی می‌پردازیم. برای این منظور، دو معماری مختلف یادگیری انتقالی، شامل دِنس-نت و اِکسپشن^۱ با تنظیمات مختلف، بررسی می‌شوند. بسته به ساختار معماری اِکسپشن، قسمت طبقه‌بندی سه کلاسه نیز دستخوش تغییرات در تعداد لایه‌ها، نورون‌ها و تنظیمات اَبَرپارامترها می‌شود. آزمایشات با تنظیمات مختلفی از جمله اندازه‌های دسته^۲، نرخ یادگیری، و فن‌های افزایش داده^۳ متفاوت انجام می‌شود تا دلیل انتخاب مدل بهینه مناسب سامانه پیشنهادی نشان داده شود. جزئیات طراحی مدل‌ها و تنظیمات مربوط به آنها در زیربخش بعدی تشریح می‌شود.

۴-۱-۱- جزئیات و تنظیمات مدل‌های یادگیری عمیق

به منظور ارزیابی و انتخاب مدل یادگیری مناسب برای سامانه

از دست داده و یا دچار مشکل شود. به همین دلیل برای جلوگیری از سوءگیری کاربر و امکان مواجهه با کلاس‌های مختلف، شرط انتخاب حداقل ۲۰ درصد از تعداد کل سؤالات در هر شرایطی از هر کلاس اعمال می‌شود.

پس از مشخص شدن تعداد سؤالات انتخابی برای هر کلاس، ترکیب درجه سختی سؤالات برای هر کلاس تعیین می‌شود. رابطه (۶) تعریف ریاضی نحوه‌ی محاسبه‌ی نرخ پاسخ‌های غلط در هر کلاس بر اساس سه درجه سختی آسان، متوسط و سخت را بیان می‌کند. در این رابطه، $Num\ Incorrect_{(c,d)}$ تعداد پاسخ‌های غلط در کلاس c با درجه سختی d است و به تبع آن $Num\ correct_{(c,d)}$ تعداد پاسخ‌های صحیح در کلاس c با درجه سختی d است. منظور از $Incorrect\ Ratio_{(c,d)}$ نیز، نرخ پاسخ‌های غلط برای سؤالات با درجه سختی d در کلاس c است.

$$Incorrect\ Ratio_{(c,d)} = \frac{Num\ Incorrect_{(c,d)}}{Num\ Incorrect_{(c,d)} + Num\ correct_{(c,d)}} \quad (6)$$

رابطه (۷) نیز تعریف ریاضی نحوه‌ی محاسبه‌ی تعداد سؤالات مربوط به هر درجه سختی در هر کلاس را بیان می‌کند. در این رابطه $N(class,difficulty)$ تعداد سؤالات انتخابی با درجه سختی d از کلاس c نشان می‌دهد.

$$N(class,difficulty) = \max(((4) \times N_{Total}, Incorrect\ Ratio_{(class,difficulty)} \times N_{class})) \quad (7)$$

مشابه آنچه در بخش تعیین تعداد کل سؤالات یک کلاس گفته شد، اگر کاربر بتواند در یک کلاس، در سؤالات با درجه سختی خاصی به تسلط بالایی برسد، در حالت عادی این مهم باعث خواهد شد تا وزن سؤالات با آن درجه سختی در کلاس مد نظر کاهش یافته و به تبع آن تعداد سؤالات انتخابی از آن کلاس به حداقل و حتی به صفر برسد. به دلیل آن که آموزش در حوزه‌ی تصاویر پزشکی، نیازمند کسب تجربه با مشاهده مداوم نمونه‌های مختلف میسر می‌شود، این روش ممکن است در چندین مرحله باعث شود کاربر تسلط به دست آمده در یک درجه سختی از کلاس مد نظر را از دست داده و یا دچار مشکل شود. به همین دلیل برای جلوگیری از سوگیری کاربر به یک درجه سختی خاص و امکان مواجهه با سؤالات با درجه سختی متنوع در هر مرحله از آزمون، شرط انتخاب حداقل ۴ درصد از تعداد کل سؤالات در هر شرایطی از هر درجه سختی برای هر کلاس اعمال

³ Data Augmentation

¹ Xception

² DenseNet

پیشنهادی، از ۱۴ مدل مختلف با پیکربندی‌های متنوع استفاده می‌شود. مدل‌ها از سه منظر بررسی می‌شوند: (۱) معماری مورد استفاده، (۲) تنظیمات نرخ یادگیری و اندازه دسته، و (۳) روش‌های

شکل ۱۰.

شکل ۱۱. جدول ۱. جزئیات و تنظیمات ۱۴ مدل آزمایش شده

شماره مدل	مدل یادگیری انتقالی	اندازه دسته	فن افزایش داده	نرخ یادگیری	شرط توقف زودهنگام
۱	دِنس-نت	۶۴	نوع ۱	۰/۰۰۱	۵ دوره
۲	دِنس-نت	۶۴	نوع ۲	۰/۰۰۱	۵ دوره
۳	دِنس-نت	۱۲۸	نوع ۱	۰/۰۰۱	۵ دوره
۴	دِنس-نت	۱۲۸	نوع ۲	۰/۰۰۱	۵ دوره
۵	دِنس-نت	۶۴	نوع ۳	۰/۰۰۱	۵ دوره
۶	دِنس-نت	۶۴	نوع ۴	۰/۰۰۱	۵ دوره
۷	دِنس-نت	۳۲	نوع ۴	۰/۰۰۳	۱۰ دوره
۸	دِنس-نت	۶۴	نوع ۴	۰/۰۰۱	۱۰ دوره
۹	دِنس-نت	۱۲۸	نوع ۴	۰/۰۰۱	۵ دوره
۱۰	دِنس-نت	۱۲۸	نوع ۳	۰/۰۰۱	۵ دوره
۱۱	اِکسپشن	۱۶	نوع ۲	۰/۰۰۱	۱۰ دوره
۱۲	اِکسپشن	۶۴	نوع ۲	۰/۰۰۱	۱۰ دوره
۱۳	اِکسپشن	۱۲۸	نوع ۲	۰/۰۰۱	۱۰ دوره
۱۴	اِکسپشن	۱۲۸	نوع ۱	۰/۰۰۱	۱۰ دوره

جدول ۲. فنون بکار گرفته شده در هر یک از انواع افزایش داده‌ی استفاده شده بر روی تصاویر پایگاه داده

نوع افزایش داده	فن			
	بزرگ‌نمایی	چرخش	انعکاس افقی	جابجایی
نوع ۱	✓	✓	X	X
نوع ۲	✓	✓	✓	X
نوع ۳	X	X	X	✓
نوع ۴	X	X	X	X

همچون نرخ یادگیری و شروط جلوگیری‌کننده از بیش‌برازش باشد،

از کیفیت داده‌ی ورودی و پیش‌پردازش‌های انجام شده است.

برخلاف تصور ما، با توجه به اینکه مجموعه‌داده استفاده شده از منابع مختلف گردآوری شده است و دارای تنوع مناسب در معیارهای مختلف از جمله تنوع در کیفیت دستگاه‌های تصویربرداری است، اعمال فن‌های افزایش داده، کمک شایانی در بهبود عملکرد مدل‌ها نداشته است. همچنان که در جدول ۳ قابل مشاهده است، کمترین میزان دقت در مدل‌های معرفی شده، به ترتیب مربوط به مدل ۴ و ۲ است که از فن‌های مختلفی برای پیش‌پردازش داده‌های ورودی در آن‌ها استفاده شده است.

۴-۱-۲- ارزیابی و مقایسه عملکرد مدل‌ها

عملکرد ۱۴ مدل معرفی شده در بخش قبلی بر اساس چهار معیار ارزیابی شامل صحت، بازیابی، امتیاز-اف-۲، و دقت ارزیابی شده و نتایج حاصل از آزمایشات در جدول ۳ ارائه شده‌اند. همان‌طور که در این جدول مشاهده می‌شود، مدل ۶ (بدون استفاده از فن‌های افزایش داده) و مدل ۵ (با استفاده از فن افزایش داده تغییر کنتراست) به ترتیب بهترین عملکرد را نسبت به ۱۲ مدل دیگر کسب کرده‌اند. مدل ۶ با دقت ۹۱/۹۳ درصد و مدل ۵ با دقت ۹۱/۸۵ درصد برترین نتیجه را بر روی داده‌های ارزیابی داشته‌اند. از نتایج ارزیابی مدل‌های مختلف می‌توان متوجه شد که تأثیرپذیری مدل بیش از آن که از اندازه دسته‌های ورودی و دیگر عامل‌های مدل

جدول ۳. عملکرد ۱۴ مدل مختلف بر اساس چهار معیار ارزیابی

شماره مدل	صحت (%)	بازیابی (%)	امتیاز-اف ۲ (%)	دقت (%)
۱	۸۶/۸۴	۸۴/۴۷	۸۴/۵۳	۸۴/۴۷
۲	۸۵/۶۹	۸۴/۲۴	۸۴/۳۲	۸۴/۲۴
۳	۸۶/۹۷	۸۵/۲۵	۸۵/۲۹	۸۵/۲۵
۴	۸۵/۷۸	۸۳/۱۵	۸۳/۲۲	۸۳/۱۵
۵	۹۲/۵۳	۹۱/۸۵	۹۱/۸۷	۹۱/۸۵
۶	۹۲/۳۸	۹۱/۹۳	۹۱/۹۵	۹۱/۹۳
۷	۹۱/۷۰	۹۱/۳۸	۹۱/۴۱	۹۱/۳۸
۸	۸۸/۹۹	۸۷/۵۸	۸۷/۶۳	۸۷/۵۸
۹	۹۰/۸۱	۸۰/۵۳	۹۰/۵۵	۹۰/۵۳
۱۰	۹۰/۱۰	۸۹/۶۰	۸۹/۶۴	۸۹/۶۰
۱۱	۸۵/۸۹	۸۹/۶۰	۸۸/۷۷	۹۰/۵۳
۱۲	۸۶/۷۵	۸۸/۶۴	۸۸/۲۱	۹۰/۷۶
۱۳	۸۸/۴۶	۸۷/۵۵	۸۷/۶۹	۹۱/۴۶
۱۴	۸۵/۵۱	۸۶/۳۶	۸۶/۰۳	۸۹/۷۵

به درستی تشخیص داده‌اند و از بیش برآزش جلوگیری کرده‌اند. بر اساس ارزیابی‌ها و بررسی‌های تشریح شده فوق، از میان ۱۴ مدل بررسی شده، مدل ۶ با دقت تشخیصی ۹۱/۹۳ درصد به‌عنوان مدل برتر انتخاب می‌شود. نتایج به‌دست‌آمده نشان می‌دهد که معماری دنس-نت در مقایسه با اکسپشن عملکرد بهتری در اکثر معیارهای کلیدی، به‌ویژه دقت نهایی و توانایی تعمیم به داده‌های جدید، دارد. علاوه بر این، معماری دنس-نت به دلیل اتصال‌های متراکم بین لایه‌ها و استفاده کارآمدتر از ویژگی‌های استخراج شده، توانایی بیشتری در یادگیری و شناسایی الگوهای پیچیده دارد. این معماری ضمن کاهش تعداد پارامترها نسبت به اکسپشن، منجر به پایداری بهتر در فرآیند آموزش و جلوگیری از نوسانات عملکرد در داده‌های اعتبارسنجی نیز می‌شود. همچنین معماری اکسپشن به دلیل پیچیدگی بیشتر و حساسیت بالا به تنظیمات آبرپارامترها، به تنظیمات دقیق‌تری نیاز دارد که منجر به افزایش بار زمانی و منابع محاسباتی می‌شود. با لحاظ کردن موارد فوق، مدل ۶ (با معماری دنس-نت) به دلیل توانایی بالا در پیش‌بینی صحیح داده‌های نادیده و عدم نیاز به اعمال افزایش داده، در سامانه پیشنهادی مورد استفاده قرار گرفت.

۴-۲- ارزیابی عملکرد سامانه آموزشی شخصی سازی شده و تحلیل عملکرد کاربران

در این بخش، نتایج حاصل از استفاده کاربران واقعی از سامانه پیشنهادی تحلیل می‌شود. در آزمایشات انجام شده تعداد ۱۰ نفر رزیدنت رادیوگرافی داوطلب مشارکت کردند. با دریافت مشاوره‌های لازم از پزشکان و اساتید دانشگاه متخصص حوزه سعی بر این بوده

در شکل ۱۰ منحنی‌های خطا و دقت داده‌های بخش آموزش و اعتبارسنجی دو مدل ۶ و ۵ که به ترتیب بالاترین میزان دقت را در بین مدل‌های معرفی شده به دست آورده‌اند، نشان داده شده است. منحنی‌های ارائه شده عملکرد مدل یادگیری عمیق در دوره‌های آموزش را به تصویر می‌کشند. این نمودارها شامل دقت و خطا برای هر دو مجموعه داده‌ی آموزش و اعتبارسنجی هستند. نتایج نشان می‌دهد که هر دو مدل با گذر زمان بهبود قابل‌توجهی داشته‌اند و این بهبود در هر دو مجموعه آموزش و اعتبارسنجی مشهود است. در ابتدا قرار بود که مدل‌ها به مدت ۱۰۰ دوره آموزش داده شود، اما با استفاده از فن Callback با دو تابع توقف زودهنگام و کاهش نرخ یادگیری، آموزش مدل به طور خودکار در دوره ۲۵ برای مدل ۶ و دوره ۱۶ برای مدل ۵ متوقف شده است. این توقف خودکار نشان‌دهنده تثبیت عملکرد مدل‌ها و جلوگیری از آموزش بیش از حد است. نکته مهم و حائز اهمیت دیگری که از این نمودارها برداشت می‌شود، دقت و خطای بهتر در مجموعه داده‌های اعتبارسنجی نسبت به داده‌های آموزش است. به طور معمول، انتظار می‌رود که دقت روی داده‌های آموزشی بهتر باشد و خطا نیز روی داده‌های اعتبارسنجی بیشتر باشد. با این حال، در این مدل‌ها نتایج خلاف این انتظار رخ داده است. این موضوع نشان می‌دهد که مدل‌ها بر روی داده‌های اعتبارسنجی بهتر عمل کرده و توانسته‌اند داده‌های نادیده را به‌خوبی پیش‌بینی کنند. همچنین، در نمودار خطا، کاهش پیوسته‌ی خطا در هر دو مدل و در هر دو مجموعه داده‌ی آموزش و اعتبارسنجی مشاهده می‌شود. این کاهش هم‌زمان خطا بیانگر آن است که مدل‌ها نه تنها توانسته‌اند داده‌های آموزشی را به‌خوبی یاد بگیرند، بلکه الگوهای موجود در داده‌های اعتبارسنجی را نیز

شرط توقف آموزش کاربر تعریف شده است (به همین دلیل در جدول ۴ تعداد آزمون‌های کاربران مختلف متفاوت است). این حد آستانه برای ارزیابی اولیه و آزمایشگاهی سامانه تعیین شده و طبیعتاً در استفاده بالینی و دانشگاهی، این مقدار به مراتب نسبت به استانداردهای تعریفی بالاتر و دقیق‌تر تنظیم می‌شود.

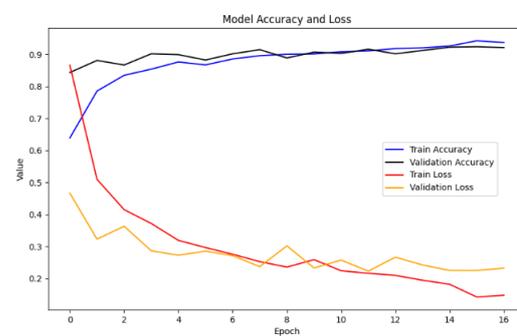
نتایج جدول ۴ نشان می‌دهد که کاربران در آزمون‌های آغازین با دقت‌های نسبتاً پایینی شروع کرده‌اند (بین ۳۹/۲۳ درصد تا ۷۰/۵۹ درصد در آزمون اول). این موضوع نشان می‌دهد که در ابتدا کاربران در تشخیص صحیح تصاویر با چالش‌های مختلفی مواجه بوده‌اند. با این حال، روند پیشرفت کاربران در طول آزمون‌های متوالی بعدی به طور واضح قابل مشاهده است. به مرور زمان و با کمک سامانه آموزشی، دقت کلی همه کاربران بهبود یافته است. به عنوان نمونه، کاربر شماره ۱ از دقت ۵۳/۰۳ درصد در آزمون اول به دقت ۹۰/۳۱ درصد در آزمون پنجم رسیده است. این نشان‌دهنده تأثیر مثبت سامانه در افزایش توانایی‌های تشخیصی کاربران است.

بسیاری از کاربران پس از چند آزمون به دقت بالای ۸۰ درصد رسیده‌اند و در برخی موارد به دقت نزدیک به ۹۶ درصد نیز دست یافته‌اند (مانند کاربران ۷، ۸ و ۱۰). این امر نشان می‌دهد که سامانه آموزشی توانسته است به خوبی کاربران را به سطحی برساند که دقت تشخیصی قابل قبولی داشته باشند.

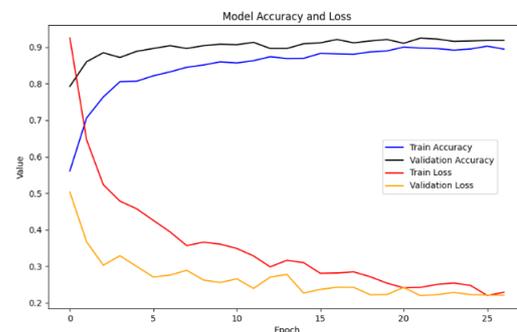
لازم به ذکر است که برخی از کاربران در برخی آزمون‌ها کاهش دقت را تجربه کرده‌اند (مانند کاربر ۳ در آزمون ۳). این نوسانات ممکن است به دلیل پیچیدگی سؤالات یا عدم تمرکز کافی کاربران در برخی آزمون‌ها باشد. بررسی جزئیات دقت هر کاربر نشان می‌دهد که سامانه آموزشی توانسته است در تمام سطوح، از کاربران ضعیف‌تر تا کاربران قوی‌تر، تأثیرات مثبت داشته باشد. کاربران ضعیف‌تر (مانند کاربران ۵ و ۶) که در ابتدا دقت‌های پایینی داشته‌اند، با ادامه استفاده از سامانه به دقت‌های بالایی دست یافته‌اند (به ترتیب ۸۸/۸۱ درصد و ۹۳/۱۲ درصد).

در نهایت، نتایج حاکی از آن است که سامانه آموزشی شخصی‌سازی شده نه تنها توانسته است دقت تشخیصی کاربران را بهبود بخشد، بلکه با ارائه محتوای آموزشی مناسب، روند یادگیری را نیز تسریع کرده است. عملکرد کاربران در آزمون‌های متوالی نشان از توانمندی سامانه در تشخیص نیازهای آموزشی کاربران و ارائه محتوای مناسب به آن‌ها دارد. این موفقیت در آزمایش‌های اولیه، نویدبخش کاربردهای وسیع‌تر سامانه در محیط‌های بالینی و دانشگاهی است.

است که حتی‌المقدور کلیه جوانب آموزشی و پزشکی در انجام آزمایشات سامانه رعایت گردد. برای ارزیابی عملکرد سامانه آموزشی و تحلیل عملکرد کاربران آن، معیارهای مختلفی نظیر دقت تشخیص کلی، دقت مثبت شده برای هر کلاس، نرخ مثبت کاذب و منفی کاذب به تفکیک کلاس‌های تعریف شده و ماتریس درهم‌ریختگی و همچنین تعداد سؤالات پاسخ داده شده صحیح و ناصحیح به تفکیک درجه سختی (آسان، متوسط و سخت) برای هر آزمون و هر کاربر مورد بررسی قرار می‌گیرند. در ادامه، ابتدا دقت کلی عملکرد کاربران سامانه ارزیابی و تحلیل می‌شود و سپس دقت عملکرد هر یک از کاربران بصورت جزئی و بر روی کلاس‌های مختلف بیماری‌ها گزارش شده و تحلیل می‌شوند.



(الف)



(ب)

شکل ۱۲. شکل ۱۰. منحنی‌های خطا و دقت بدست آمده روی مجموعه داده‌ی آموزشی و اعتبارسنجی در طی تکرارهای آموزشی توسط برترین مدل‌ها. (الف) مدل ۶ و (ب) مدل ۵

۴-۲-۱- ارزیابی دقت کلی عملکرد کاربران

در جدول ۴ دقت کلی به دست آمده توسط ۱۰ کاربر سامانه طی ۶ آزمون متوالی گزارش شده است. در برگزاری آزمون‌ها، حد آستانه دستیابی به دقت بیش از ۸۰ درصد در تمامی سه کلاس به عنوان

جدول ۴. دقت کلی (بر اساس درصد) به‌دست‌آمده توسط ۱۰ کاربر در آزمون‌های برگزار شده

شماره کاربر	شماره آزمون				
	۱	۲	۳	۴	۵
۱	۵۳/۰۳	۵۹/۵۳	۵۹/۵۶	۷۸/۰۸	۹۰/۳۱
۲	۵۰/۶۲	۶۹/۱۸	۹۰/۱۳	-	-
۳	۵۰/۳۶	۶۳/۲۳	۵۳/۳۵	۶۸/۳۸	۷۷/۲۹
۴	۵۳/۱۵	۴۵/۷۷	۷۳/۰۲	۶۹/۹۳	۸۳/۹۶
۵	۴۶/۷۶	۷۱/۱۴	۸۸/۸۱	-	-
۶	۳۹/۲۳	۵۵/۰۸	۸۱/۱۲	۹۳/۱۲	-
۷	۶۱/۴۹	۶۲/۹۱	۸۱/۱۵	۹۶/۱۵	-
۸	۷۰/۵۹	۸۶/۳۰	۹۶/۵۹	-	-
۹	۶۴/۹۶	۸۱/۶۳	۸۶/۹۲	-	-
۱۰	۶۰/۱۸	۶۸/۱۶	۸۲/۱۵	۹۵/۲۶	-

در کلاس ذات‌الرینه، دقت این کاربر نوسانات بیشتری داشته است. دقت از ۵۰٪ در آزمون اول به ۲۲٪ در آزمون دوم کاهش یافته، سپس در آزمون‌های بعدی به تدریج افزایش یافته و در آزمون پنجم به ۸۲٪ رسیده است. با بررسی نتایج حاصل توسط این کاربر، می‌توان گفت این نوسانات ممکن است به دلیل افزایش حساسیت کاربر نسبت به سؤالات با پیچیدگی بالا و غفلت از مسائل ساده‌تر که ناشی از نتایج آزمون اول است رخ داده باشد. در ادامه با افزایش داده‌ها و شخصی‌سازی شدن سؤالات بر اساس مدل پیشنهادی، کاربر ۱ توانسته است تشخیص صحیح سؤالات با پیچیدگی کم‌تر را نیز بهبود دهد. همچنین ممکن است این نوسانات، ناشی از پیچیدگی تشخیص ذات‌الرینه و یا نیاز به زمان بیشتر برای بهبود تشخیص در این کلاس برای این کاربر باشد.

نرخ خطای این کاربر، در آزمون اول نسبتاً بالا (۰/۷۵) بوده است. این میزان خطا نشان دهنده‌ی مشکلاتی در تشخیص صحیح موارد ذات‌الرینه بوده است. با تکرار آزمون‌ها، نرخ خطا در این کلاس نیز به طور پیوسته کاهش یافته و در آزمون پنجم به ۰/۲۴ رسیده است. این کاهش نرخ خطا نشان دهنده‌ی بهبود تدریجی در توانایی کاربر در تشخیص این کلاس است؛ هرچند که روند کاهش خطا در این کلاس نسبت به کلاس کوید-۱۹ کندتر بوده است.

این کاربر در تشخیص کلاس سالم نیز در ابتدا با چالش‌هایی روبرو بوده است، به طوری که نرخ خطا در آزمون اول برابر با ۰/۷۸ بوده است. اما به مرور زمان و با استفاده از سامانه آموزشی، نرخ خطا در این کلاس نیز کاهش یافته و در آزمون پنجم به ۰/۱۱ رسیده است. دقت کاربر در این کلاس از ۰/۳۶ در آزمون اول به ۰/۸۸ در آزمون پنجم رسیده است. در آزمون اول، درصد پاسخ‌های نادرست به سؤالات آسان‌تر (۰/۸۰) بیشتر از سؤالات سخت‌تر (۰/۵۵) بوده است. این امر ممکن است به دلیل توجه کمتر کاربر به جزئیات در

۴-۲-۲- ارزیابی و تحلیل جزئی عملکرد نمونه موردی از کاربران

در این بخش جزئیات نتایج آموزشی و تشخیصی به دست آمده توسط یک کاربر موردی (مشخصاً کاربر شماره ۱) به صورت تحلیلی و بر روی هر یک از کلاس‌های تشخیصی گزارش شده و مورد بررسی می‌گیرد.

در آزمون اول، دقت به‌دست‌آمده توسط این کاربر در کلاس کوید-۱۹، ۷۲ درصد بوده که در طی پنج آزمون طراحی شده با رشد متوالی به ۱۰۰ درصد رسیده است. همچنین مجموع نرخ خطا برای کلاس کوید-۱۹ برابر با ۰/۵۷ بوده که نشان‌دهنده وجود چالش‌های قابل توجهی برای این کاربر در تشخیص موارد کوید-۱۹ است. اما با گذشت زمان، این نرخ به تدریج کاهش یافته و در آزمون پنجم به ۰/۰۹ رسیده است. با وجود دستیابی به دقت ۱۰۰ درصد در این کلاس، همچنان نرخ خطا، مقدار مثبتی را که ناشی از خطای مثبت کاذب برای این کلاس است، نشان می‌دهد. دلیل انتخاب این معیار برای بررسی عملکرد کاربر و استفاده از آن در طراحی آزمون‌ها، بروز چنین مسائلی است. خطای مثبت کاذب در این کلاس بدین معنا است که این کاربر، به تعدادی سؤال که دارای بیماری دیگر یا سالم بوده‌اند، تشخیص کوید-۱۹ داده است که در عمل می‌تواند هزینه‌زا برای تمامی ارکان باشد.

از جنبه‌ی درجه سختی سؤالات پاسخ داده شده توسط کاربر ۱، در آزمون اول، درصد پاسخ‌های نادرست به سؤالات سخت‌تر بیشتر از سؤالات آسان‌تر بوده است. این روند نشان می‌دهد که کاربر در ابتدا با سؤالات سخت‌تر مشکل بیشتری داشته است. با گذر زمان، این درصد به تدریج کاهش یافته و در آزمون پنجم به مقدار صفر برای تمامی درجات سختی رسیده است.

بهبود عملکرد تشخیصی و کاهش خطاهای پزشکی مورد استفاده قرار گیرد.

در سامانه‌ی پیشنهادی، برای کاهش تأثیرات نامطلوب و گاه غیرقابل جبران حاصل از تشخیص‌های منفی کاذب و مثبت کاذب بر روند درمان بیماری و سیستم درمان، ابتدا یک راهکار مبتنی بر یادگیری انتقالی عمیق جهت تشخیص سه کلاس بیماری کوید-۱۹، ذات‌الریه و وضعیت سالم با استفاده از تصاویر رادیوگرافی قفسه سینه و استخراج درجه سختی تصاویر معرفی شد. در روش پیشنهادی، یکی از اهداف پیگیری این موضوع بود که مدل تولید شده، بتواند با حداقل پیش‌پردازش بر روی داده‌های ورودی به نتایج مطلوبی برسد. چراکه معمولاً کاربران سامانه، تشخیص‌های خود را با استفاده از تصاویر خام انجام می‌دهند. با استفاده از یک شبکه یادگیری انتقالی عمیق (دِنس-نت) ویژگی درجه سختی تشخیص تصاویر ورودی استخراج شد. همچنین با استفاده از یک الگوریتم تطبیقی و تعاملی طراحی شده برای انتخاب سؤالات در هر آزمون، به تدریج سطح دشواری سؤالات متناسب با عملکرد هر کاربر تنظیم شد. این روش نه تنها به افزایش دقت رزیدنت‌ها در تشخیص بیماری‌های مرتبط با تصاویر رادیوگرافی قفسه سینه کمک می‌کند، بلکه با تمرکز بر خطاهای تشخیصی گذشته، فرایند یادگیری را برای هر کاربر به صورت شخصی‌سازی شده ارائه می‌دهد.

نتایج حاصل از بررسی کارایی سامانه بر اساس آزمون‌های واقعی نشان داد که کاربران با تکرار آزمون‌ها و دریافت بازخوردهای تعاملی، به تدریج دقت خود را در تشخیص بیماری‌ها بهبود داده و میزان خطاهای مثبت کاذب و منفی کاذب خود را کاهش می‌دهند. همچنین، تحلیل دقیق نتایج آزمون‌ها به تفکیک هر کلاس و درجه سختی سؤالات، نشان دهنده تأثیر مثبت این رویکرد آموزشی بر یادگیری و توانمندسازی رزیدنت‌ها و کارکنان کادر درمان بود.

نوآوری بارز این پژوهش در تلفیق یکپارچه‌ی یادگیری عمیق، آزمون‌های تطبیقی و تحلیل عملکرد کاربران در قالب یک سامانه‌ی شخصی‌سازی‌شده‌ی آموزش برخط نهفته است. برخلاف پژوهش‌های پیشین [۸] تا [۱۰] و [۱۲] تا [۱۸] که عمدتاً فقط محدود به تحلیل خطاهای تشخیصی مثبت کاذب یا فقط منفی کاذب بودند، رویکرد ارائه‌شده در این پژوهش با تمرکز بر طراحی آزمون‌های پویای مبتنی بر تحلیل عملکرد پیشین کاربران با استفاده از تصاویر رادیوگرافی قفسه سینه و ارائه بازخورد دقیق از نقاط ضعف در خطاهای مثبت کاذب و منفی کاذب بصورت همزمان، بستری عملی و قابل اجرا در محیط بالینی برای آموزش رادیولوژیست‌ها فراهم ساخته است. همچنین، استفاده از مدل‌های عمیق برای

سؤالات آسان‌تر باشد. با گذر زمان، درصد پاسخ‌های نادرست به سؤالات آسان به تدریج کاهش یافته و در آزمون پنجم، به مقدار قابل قبولی (۰/۰۸) رسیده است.

در جمع‌بندی برای کاربر ۱ می‌توان گفت، کاربر در تشخیص موارد کوید-۱۹ به تسلط کامل رسیده است و هیچ خطایی در آزمون‌های نهایی نداشته است ولیکن همچنان با خطای مثبت کاذب درگیر است که نیازمند چند مرحله آموزش بیشتر است. در مورد ذات‌الریه این کاربر با وجود پیشرفت، همچنان در تشخیص موارد ذات‌الریه چالش‌هایی داشته است که این امر نشان‌دهنده‌ی پیچیدگی بیشتر این کلاس برای این کاربر است. در تشخیص موارد سالم نیز پیشرفت قابل توجهی داشته، اما همچنان نیاز به بهبود در مواجهه با سؤالات با درجات سختی متفاوت دارد. این تحلیل نشان دهنده‌ی تأثیر مثبت سامانه آموزشی بر بهبود عملکرد تشخیصی این کاربر است و همچنین به نقاطی که کاربر نیاز به تمرین بیشتر دارد، اشاره می‌کند. با وجود آن‌که در این بخش تنها عملکرد یک کاربر به صورت تفصیلی بررسی شد، مرور نتایج سایر کاربران نشان می‌دهد که الگوی کلی پیشرفت تقریباً در تمامی موارد مشابه و رو به بهبود بوده است. اغلب کاربران در ابتدا دارای سطح دقت متوسط یا پایین بوده‌اند، اما در طول مراحل آزمون و با بهره‌گیری از سامانه، دقت و صحت عملکرد آن‌ها به طور پیوسته بهبود یافته و نرخ خطاهای مثبت و منفی کاذب کاهش یافته است.

تفاوت‌های اصلی در میان کاربران بیشتر به سرعت یادگیری و توانایی مقابله با سؤالات پیچیده‌تر مربوط می‌شود. برخی افراد در مدت کوتاهی به سطح بالایی از عملکرد دست یافته‌اند، در حالی که برخی دیگر مسیر تدریجی‌تری را طی کرده‌اند و نیازمند تمرین و بازخورد بیشتری بوده‌اند. در مجموع، نتایج به دست آمده از تمامی نمونه‌ها مؤید آن است که سامانه پیشنهادی در ارتقای مهارت‌ها و دقت تشخیصی کاربران مؤثر بوده و توانسته است روندی باثبات از بهبود عملکرد را ایجاد کند.

۵- نتیجه‌گیری و کارهای آینده

در این پژوهش، به طراحی و پیاده‌سازی یک سامانه آموزشی آنلاین برای شخصی‌سازی آموزش رزیدنت‌های رادیولوژی با تمرکز بر استخراج تعاملی مدل خطاهای تشخیصی مثبت کاذب و منفی کاذب در تصاویر رادیوگرافی قفسه سینه پرداخته شد. هدف اصلی این سامانه، افزایش دقت تشخیصی رزیدنت‌ها از طریق آزمون‌های متوالی و تحلیل خطاها بود. باتوجه به اهمیت دقت در تشخیص‌های پزشکی، مخصوصاً از روی تصاویر رادیولوژی، و نیاز به آموزش مداوم و بهینه برای رزیدنت‌ها، این سامانه می‌تواند به عنوان ابزاری مؤثر در

- [5] B. Qi et al., GREN: Graph-Regularized Embedding Network for Weakly-Supervised Disease Localization in X-ray images. 2021.
- [6] V. Sorin, Y. Barash, E. Konen, and E. Klang, "Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review," *Journal of the American College of Radiology*, vol. 17, no. 5, pp. 639-648, 2020/05/01/ 2020, doi: <https://doi.org/10.1016/j.jacr.2019.12.026>.
- [7] L. Guo, A. Tahir, D. Zhang, Z. Wang, and R. Ward, *Automatic Medical Report Generation: Methods and Applications*. 2024.
- [8] M. Mazurowski, J. Baker, H. Barnhart, and G. Tourassi, "Individualized computer-aided education in mammography based on user modeling: Concept and preliminary experiments," *Medical physics*, vol. 37, pp. 1152-60, 03/01 2010, doi: 10.1118/1.3301575.
- [9] M. Wang, M. Wang, L. J. Grimm, and M. A. Mazurowski, "A computer vision-based algorithm to predict false positive errors in radiology trainees when interpreting digital breast tomosynthesis cases," *Expert Systems with Applications*, vol. 64, pp. 490-499, 2016/12/01/ 2016, doi: <https://doi.org/10.1016/j.eswa.2016.08.023>.
- [10] M. Wang et al., "Predicting false negative errors in digital breast tomosynthesis among radiology trainees using a computer vision-based approach," *Expert Systems with Applications*, vol. 56, pp. 1-8, 2016/09 /۲۰۱۶ /۰۱ /doi: <https://doi.org/10.1016/j.eswa.2016.01.053>.
- [11] S. Burti, A. Zotti, and T. Banzato, "Role of AI in diagnostic imaging error reduction," *Frontiers in Veterinary Science*, vol. 11, 08/30 2024, doi: 10.3389/fvets.2024.1437284.
- [12] M. A. Mazurowski and G. D. Tourassi, "Exploring the potential of collaborative filtering for user-adaptive mammography education," in *Proceedings of the 2011 Biomedical Sciences and Engineering Conference: Image Informatics and Analytics in Biomedicine*, 15-17 March ۲۰۱۱ /۲۰۱۱ pp. 1-4, doi: 10.1109/BSEC.2011.5872325 .
- [13] S. Voisin, F. Pinto, G. Morin-Ducote, K. B. Hudson, and G. D. Tourassi, "Predicting diagnostic error in radiology via eye-tracking and image analytics: preliminary investigation in mammography," (in eng), *Med Phys*, vol. 40, no. 10, p. 101906, Oct 2013, doi: 10.1118/1.4820536.
- [14] H. Lin, X. Yang, and W. Wang, "A content-boosted collaborative filtering algorithm for personalized training in interpretation of radiological imaging," *Journal of digital imaging*, vol. 27, pp. 449-456, 2014.
- [15] M. A. Mazurowski, J. Zhang, J. Y. Lo, C. M. Kuzmiak, S. V. Ghate, and S. Yoon, "Modeling resident error-making patterns in detection of mammographic masses using computer-extracted image features: preliminary experiments," in *Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*, 2014, vol. 9037: SPIE, pp. 197-202 .
- [16] J. Zhang, J. Y. Lo, C. M. Kuzmiak, S. V. Ghate, S. C. Yoon, and M. A. Mazurowski, "Using computer- extracted image features for modeling of error- making patterns in detection of mammographic masses among radiology residents," *Medical physics*, vol. 41, no. 9, p. 091907, 2014.
- [17] J. Zhang, J. I. Silber, and M. A. Mazurowski, "Modeling false positive error making patterns in radiology trainees for improved mammography education," *Journal of biomedical informatics*, vol. 54, pp. 50-57, 2015.
- [18] Z. Gandomkar, K. Tay, W. Ryder, P. Brennan, and C. Mello-Thoms, Predicting radiologists' true and false positive decisions in reading خودکارسازی تحلیل‌ها، این سامانه را نسبت به سیستم‌های آموزشی گذشته کاربردی‌تر و مقیاس‌پذیرتر کرده است.
- از محورهای مطالعه و توسعه بیشتر این پژوهش می‌توان به کاربرد سامانه پیشنهادی در سایر تصاویر پزشکی و بیماری‌های مرتبط اشاره کرد. این امر موجب تنوع بیشتر در آموزش و افزایش قابلیت‌های سامانه خواهد شد. همچنین با بهره‌گیری از الگوریتم‌های یادگیری عمیق و شبکه‌های عصبی کانولوشنی دقیق‌تر و بیشتر و استفاده از روش‌های یادگیری جمعی و گروهی^۱ می‌توان دقت در تشخیص الگوی خطاهای مثبت کاذب و منفی کاذب را با دستیابی به معیارهای دقیق‌تر و قابل‌تعمیم‌تر تعیین‌کننده‌ی درجه سختی تصاویر بهبود بخشید. این الگوریتم‌ها می‌توانند با تحلیل الگوهای پیچیده‌تر در تصاویر، دقت بیشتری را در سامانه ایجاد کنند. طراحی و پیاده‌سازی ویژگی‌های تعاملی بیشتر مانند بازخوردهای بصری دقیق‌تر و نقشه‌های حرارتی برای نمایش نواحی مشکوک در تصاویر، می‌تواند به فهم بهتر رزیدنت‌ها از نقاط ضعف و قوت خود در تشخیص کمک کند و یادگیری را کارآمدتر سازد. همچنین استفاده از فن‌های قطعه‌بندی برای نمایش نواحی دارای درگیری به بیماری در تصاویر مورد بررسی، در مراحل آموزش کاربر می‌تواند در یافتن نواحی و موارد تشخیص داده نشده کمک بسزایی بکند. پیشنهاد می‌شود مطالعه‌ای طولانی مدت برای بررسی تأثیرات ماندگار این نوع آموزش بر دقت تشخیصی رزیدنت‌ها انجام شود. این مطالعه می‌تواند نشان دهد که آیا آموزش شخصی‌سازی شده موجب بهبود مداوم در عملکرد تشخیصی می‌شود یا نیاز به بازبینی و اصلاح دوره‌ای دارد.

مراجع

- [1] A. P. Brady, "Error and discrepancy in radiology: inevitable or avoidable?," (in eng), *Insights Imaging*, vol. 8, no. 1, pp. 171-182, 2017, doi: 10.1007/s13244-016-0534-1.
- [2] S. Waite et al., "A Review of Perceptual Expertise in Radiology-How it develops, How we can test it, and Why humans still matter in the era of Artificial Intelligence," *Academic Radiology*, vol. 27, no. 1, pp. 26-38, 2020/01/01/ 2020, doi: <https://doi.org/10.1016/j.acra.2019.08.018>.
- [3] J. Irvin et al., "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 590-597, 07/17 2019, doi: 10.1609/aaai.v33i01.3301590.
- [4] M. Perumal, A. Nayak, R. P. Sree, and M. Srinivas, "INASNET: Automatic identification of coronavirus disease (COVID-19) based on chest X-ray using deep neural network," *ISA Transactions*, 2022/03/03/ 2022, doi: <https://doi.org/10.1016/j.isatra.2022.02.033>.

¹ Ensemble Learning

- [28] J. Cohen, P. Morrison, L. Dao, K. Roth, T. Duong, and M. Ghassemi, *COVID-19 Image Data Collection: Prospective Predictions Are the Future*. 2020.
- [29] A. Chung, "Actualmed COVID-19 chest x-ray data initiative," 2020. [Online]. Available: <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>.
- [30] A. Chung, "Figure 1 COVID-19 chest x-ray data initiative," 2020. [Online]. Available: <https://github.com/agchung/Figure1-COVID-chestxray-dataset>.
- [31] P. Garg, M. Gautam, B. Chugh, and K. Dwivedi, "Employing transfer learning techniques for COVID-19 detection using chest X-ray," *International Journal of Advances in Applied Sciences*, vol. 13, p. 680, 09/01 2024, doi: 10.11591/ijaas.v13.i3.pp680-688.
- [32] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21-26 July 2017 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243 .
- [33] J. Yu *et al.*, "A Unified Adaptive Testing System Enabled by Hierarchical Structure Search," in *Forty-first International Conference on Machine Learning* .
- [34] Q. Liu *et al.*, "Survey of computerized Adaptive Testing: A machine learning perspective," 2024 2024, doi: 10.48550/ARXIV.2404.00712.
- [35] S. J. Chen, A. Choi, and A. Darwiche, "Computer Adaptive Testing Using the Same-Decision Probability," in *BMA@ UAI*, 2015, pp. 34-43 .
- [36] J.-J. Vie, F. Popineau, É. Bruillard, and Y. Bourda, "A Review of Recent Advances in Adaptive Assessment," in *Learning Analytics: Fundamentals, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning*, A. Peña-Ayala Ed. Cham: Springer International Publishing, 2017, pp. 113-142.
- mammograms by using gaze parameters and image-based features. 2016, p. 978715.
- [19] B. Ibragimov and C. Mello-Thoms, "The Use of Machine Learning in Eye Tracking Studies in Medical Imaging: A Review," *IEEE journal of biomedical and health informatics*, vol. PP, 02/29 2024, doi: 10.1109/JBHI.2024.3371893.
- [20] H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, and R. J. Mislevy, *Computerized Adaptive Testing: A Primer*. Taylor & Francis, 2000.
- [21] W. A. Sands, B. K. Waters, and J. R. McBride, *Computerized adaptive testing: From inquiry to operation*. American Psychological Association, 1997.
- [22] E. E. Roskam and P. G. Jansen, "A new derivation of the Rasch model," in *Advances in Psychology*, vol. 20: Elsevier, 1984, pp. 293-307.
- [23] A. D. Mead" ,An introduction to multistage testing," *Applied Measurement in Education*, vol. 19, no. 3, pp. 185-187, 2006.
- [24] Y. Zhuang, Q. Liu, Z. Huang, Z. Li, S. Shen, and H. Ma, "Fully adaptive framework: Neural computerized adaptive testing for online education," in *Proceedings of the AAAI conference on artificial intelligence*, 2022, vol. 36, no. 4, pp. 4734-4742 .
- [25] D. F. Mujtaba and N. R. Mahapatra, "Artificial intelligence in computerized adaptive testing," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2020: IEEE, pp. 649-654 .
- [26] D. Kermany. Large dataset of labeled optical coherence tomography (OCT) and Chest X-Ray images, Mendeley, 2018, doi: 10.17632/RSCBJBR9SJ.3.
- [27] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, 2018/2 2018, doi: 10.1016/j.cell.2018.02.010.