دانشگاه صنعتی امیرکبیر
Amirkabir University
of Technology

تشخیص و مکان‌یابی شکستگی مچ دست کودکان در تصاویر رادیوگرافی با استفاده از YOLO۲۶

گزارش نهایی پروژه کارشناسی

دانشکده مهندسی کامپیوتر

دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران)

عنوان پروژه

تشخیص و مکان‌یابی شکستگی مچ دست کودکان در تصاویر رادیوگرافی با استفاده از YOLO۲۶

عنوان انگلیسی

Detection and Localization of Pediatric Wrist Fractures in Radiographs Using YOLO۲۶

استاد راهنما

جناب آقای دکتر علیرضا باقری

نگارش

محمد عرفان صدیق میرزایی

شماره دانشجویی: ۹۹۳۱۰۳۴

نیمسال دوم سال تحصیلی ۱۴۰۳–۱۴۰۴

تقدیر و تشکر

بدین وسیله مراتب سپاس و قدردانی خود را از جناب آقای دکتر علیرضا باقری، استاد راهنمای این پروژه، به پاس راهنمایی‌های ارزشمند، پیشنهادهای سازنده و حمایت‌های علمی ایشان در تمامی مراحل انجام این پژوهش ابراز می‌دارم. همچنین از اعضای هیئت علمی دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر که در طول دوران تحصیل زمینه رشد علمی و پژوهشی اینجانب را فراهم نمودند، صمیمانه سپاسگزاری می‌کنم. در پایان، از خانواده عزیزم که همواره با صبوری، حمایت و دلگرمی خود در تمامی مراحل تحصیل همراه من بوده‌اند، قدردانی ویژه دارم.

چکیده

تشخیص دقیق شکستگی‌های مچ دست در کودکان یکی از مهم‌ترین مسائل در تصویربرداری پزشکی و اورژانس‌های ارتوپدی به شمار می‌رود. اگرچه رادیوگرافی نخستین روش تشخیصی برای ارزیابی این نوع آسیب‌ها است، اما وجود صفحات رشد، هم‌پوشانی ساختارهای استخوانی، کیفیت متغیر تصاویر و ظرافت برخی شکستگی‌ها، فرآیند تشخیص را حتی برای پزشکان با تجربه نیز با چالش مواجه می‌کند. در سال‌های اخیر، پیشرفت روش‌های یادگیری عمیق و به‌ویژه شبکه‌های آشکارسازی اشیا، زمینه توسعه سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی را فراهم کرده است.

در این پروژه، عملکرد مدل یولو ۲۶ به‌عنوان جدیدترین نسل خانواده یولو برای تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان مورد بررسی قرار گرفته است. آموزش و ارزیابی مدل بر روی مجموعه‌داده عمومی GRAZPEDWRI-DXانجام شده است که شامل بیش از بیست هزار تصویر رادیوگرافی حاشیه‌نویسی‌شده از بیماران کودک است و یکی از معتبرترین مجموعه‌داده‌های موجود در این حوزه محسوب می‌شود.

در این پژوهش، پس از آماده‌سازی داده‌ها و انجام پیش‌پردازش‌های لازم، مدل با استفاده از تنظیمات مناسب آموزش داده شد و عملکرد آن با معیارهای متداول ارزیابی شامل دقت مثبت، فراخوانی و میانگین دقت (mAP) مورد بررسی قرار گرفت. علاوه بر تحلیل کمی، خروجی‌های مدل نیز از نظر کیفیت مکان‌یابی شکستگی‌ها و قابلیت استفاده در کاربردهای بالینی تحلیل شدند.

نتایج حاصل نشان می‌دهد که استفاده از معماری‌های جدید خانواده یولو می‌تواند دقت و سرعت بالایی را در آشکارسازی شکستگی‌های مچ دست کودکان فراهم کند و به‌عنوان یک ابزار کمک‌تشخیصی در کنار متخصصان رادیولوژی مورد استفاده قرار گیرد. همچنین این پژوهش بستری برای ارزیابی عملی نسل جدید مدل‌های آشکارسازی شیء در کاربردهای تصویربرداری پزشکی فراهم می‌آورد.

واژه‌های کلیدی

یادگیری عمیق، بینایی ماشین، آشکارسازی اشیا، یولو۲۶ ، رادیوگرافی، شکستگی مچ دست، تصویربرداری پزشکی، GRAZPEDWRI-DX

Abstract

Accurate detection of pediatric wrist fractures is one of the major challenges in musculoskeletal radiology and emergency medicine. Although radiography is the primary imaging modality for diagnosing wrist injuries, anatomical complexity, growth plates, image quality variations, and subtle fracture patterns often make image interpretation difficult, even for experienced clinicians. Recent advances in deep learning, particularly in real-time object detection, have enabled the development of intelligent computer-aided diagnosis systems capable of assisting radiologists in medical image interpretation.

This project investigates the performance of YOLO٢٦, the latest generation of the Yolo object detection family, for automatic detection and localization of pediatric wrist fractures in radiographic images. Model training and evaluation were performed using the publicly available GRAZPEDWRI-DX dataset, which contains more than twenty thousand expert-annotated pediatric wrist radiographs and has become a widely used benchmark for fracture detection research.

The proposed workflow includes dataset preparation, image preprocessing, model training, validation, and quantitative evaluation using standard object detection metrics such as Precision, Recall, and Mean Average Precision (mAP). In addition to quantitative analysis, qualitative evaluation of the predicted bounding boxes was performed to assess localization performance in clinically relevant scenarios.

The experimental results demonstrate that modern YOLO-based architectures can achieve high detection accuracy while maintaining real-time inference capability. These findings indicate that YOLO٢٦ has significant potential as a computer-aided diagnostic tool for assisting clinicians in the interpretation of pediatric wrist radiographs and provide further evidence for the applicability of state-of-the-art object detection models in medical imaging.

Keywords: Deep Learning, Computer Vision, Object Detection, YOLO٢٦, Pediatric Wrist Fracture, Medical Imaging, Radiography, GRAZPEDWRI-DX

فهرست

چکیده ۳

فصل اول کلیات پژوهش ۹

۱-۱ مقدمه ۹

۱-۲ بیان مسئله ۱۰

۱-۳ اهمیت پژوهش ۱۲

۱-۴ اهداف پژوهش ۱۳

۱-۵ سؤالات پژوهش ۱٤

۱-۶ دستاوردهای پژوهش ۱٤

۱-۷ ساختار گزارش ۱٥

فصل دوم مبانی نظری و مرور پژوهش‌ها ۱٦

۲-۱ مقدمه ۱٦

۲-۲ شکستگی‌های مچ دست در کودکان ۱۷

۲-۳ تصویربرداری رادیوگرافی در تشخیص شکستگی‌های مچ دست ۱۸

۲-۴ یادگیری عمیق و کاربرد آن در تحلیل تصاویر پزشکی ۱۹

۲-۵ آشکارسازی اشیا در تحلیل تصاویر پزشکی ۲۱

۲-۶ تکامل خانواده مدل‌های یولو و معرفی مدل یولو ۲۶ ..... ۲۲

۲-۷ مجموعه‌داده GRAZPEDWRI-DX ۲٤

۲-۸ معیارهای ارزیابی عملکرد مدل ۲٦

۲-۸-۱ مفاهیم پایه ارزیابی ۲٦

۲-۸-۳ معیار دقت مثبت ۲۸

۲-۸-۴ معیار فراخوانی ۲۸

۲-۸-۵ معیار F۱-Score ۲۸

۶-۸-۲ میانگین دقت ۲۹
۷-۸-۲ میانگین دقت مثبت(mAP) ۲۹
۸-۸-۲ جمع‌بندی ۳۰
۹-۲ پژوهش‌های پیشین ۳۰
۱-۹-۲ شکل‌گیری بستر استاندارد پژوهش ۳۰
۲-۹-۲ ارزیابی و بهینه‌سازی YOLOv۷ ..... ۳۱
۳-۹-۲ استفاده از YOLOv۸ و راهبردهای افزایش داده ۳۱
۴-۹-۲ مقایسه آشکارسازهای تک‌مرحله‌ای و دومرحله‌ای ۳۲
۵-۹-۲ به‌کارگیری یولو ۹ و یولو ۱۰ ..... ۳۲
۶-۹-۲ نسل‌های جدیدتر و تمرکز بر تعمیم‌پذیری ۳۳
۷-۹-۲ جمع‌بندی تحلیلی ۳٤
۱۰-۲ جمع‌بندی فصل و تبیین خلأ پژوهشی ۳۵
فصل سوم روش اجرای پژوهش ۳۷
۱-۳ مقدمه ۳۷
۲-۳ طراحی پژوهش و پروتکل اجرای آزمایش ۳۷
۳-۳ مجموعه‌داده مورد استفاده ۳۹
۴-۳ آماده‌سازی و پیش‌پردازش داده‌ها ٤۰
۵-۳ مدل‌های مورد بررسی ٤۲
۶-۳ محیط اجرای آزمایش و تنظیمات آموزش ٤۳
۱-۶-۳ محیط اجرای آزمایش ٤٤
۲-۶-۳ تنظیمات آموزش ٤٤
۷-۳ روش ارزیابی عملکرد مدل‌ها ٤۵
۸-۳ مراحل اجرای آزمایش ٤٦

۹-۳ جمع‌بندی فصل ٤٨
فصل چهارم نتایج ٤٩
۱-۴ مقدمه ٤٩
۲-۴ مشخصات داده‌های مورد استفاده در ارزیابی ٤٩
۳-۴ نتایج آموزش مدل یولو ۹ ..... ٥١
۱-۳-۴ روند آموزش و همگرایی مدل ٥١
۲-۳-۴ منحنی دقت مثبت–فراخوانی ٥٣
۳-۳-۴ منحنی تغییرات دقت مثبت بر حسب آستانه اطمینان ٥٤
۴-۳-۴ منحنی تغییرات فراخوانی بر حسب آستانه اطمینان ٥٥
۵-۳-۴ منحنی ضریب F۱ بر حسب آستانه اطمینان ٥٥
۶-۳-۴ ماتریس درهم‌ریختگی ٥٧
۷-۳-۴ جمع‌بندی عملکرد مدل ٥٨
۴-۴ نتایج آموزش مدل یولو ۲۶ ..... ٥٨
۱-۴-۴ روند آموزش و همگرایی مدل ٥٩
۲-۴-۴ منحنی دقت مثبت–فراخوانی ٦٠
۳-۴-۴ منحنی تغییرات دقت مثبت بر حسب آستانه اطمینان ٦١
۴-۴-۴ منحنی تغییرات فراخوانی بر حسب آستانه اطمینان ٦١
۵-۴-۴ منحنی ضریب F۱ بر حسب آستانه اطمینان ٦٢
۶-۴-۴ ماتریس درهم‌ریختگی ٦٣
۷-۴-۴ جمع‌بندی عملکرد مدل ٦٥
۵-۴ مقایسه عملکرد مدل‌های مورد بررسی ٦٥
۶-۴ جمع‌بندی فصل ٦٦
فصل پنجم جمع‌بندی، نتیجه‌گیری و پیشنهادها ٦٧

۵-۱ مقدمه ٦٧
۵-۲ جمع‌بندی پژوهش ٦٧
۵-۳ نتیجه‌گیری ٦٨
۵-۴ دستاوردهای پژوهش و کاربردهای عملی ٦٩
۵-۵ محدودیت‌های پژوهش ٧٠
۵-۶ پیشنهادهایی برای پژوهش‌های آینده ٧١
منابع ٧٢

فصل اول کلیات پژوهش

۱-۱ مقدمه

آسیب‌های اسکلتی از شایع‌ترین دلایل مراجعه کودکان به مراکز درمانی و بخش‌های اورژانس هستند و شکستگی‌های ناحیه مچ دست سهم قابل‌توجهی از این آسیب‌ها را تشکیل می‌دهند. تشخیص صحیح این آسیب‌ها در نخستین مراجعه، نقش تعیین‌کننده‌ای در انتخاب روش درمان و پیشگیری از بروز عوارض بلندمدت دارد. از آنجا که استخوان‌های کودکان در مرحله رشد قرار دارند، هرگونه خطا یا تأخیر در تشخیص می‌تواند بر روند طبیعی رشد استخوان، عملکرد اندام و کیفیت زندگی بیمار تأثیر بگذارد. ازاین‌رو، تشخیص دقیق شکستگی‌های مچ دست در کودکان از اهمیت ویژه‌ای در فرآیند تصمیم‌گیری بالینی برخوردار است[۱].

رادیوگرافی نخستین و رایج‌ترین روش تصویربرداری برای ارزیابی شکستگی‌های استخوانی به شمار می‌رود و در اغلب مراکز درمانی به‌عنوان ابزار اولیه بررسی آسیب‌های اسکلتی مورد استفاده قرار می‌گیرد[۱]. این روش با وجود مزایایی نظیر دسترس‌پذیری، سرعت انجام و هزینه نسبتاً پایین، در تفسیر تصاویر مربوط به کودکان با چالش‌های متعددی همراه است. وجود صفحات رشد، تغییرات طبیعی وابسته به سن، هم‌پوشانی ساختارهای استخوانی و ظرافت برخی شکستگی‌ها سبب می‌شود که تمایز میان ساختار طبیعی و ضایعات استخوانی در بسیاری از موارد دشوار باشد. در نتیجه، تفسیر صحیح این تصاویر به میزان قابل‌توجهی به تجربه و مهارت متخصص وابسته است و حتی در مراکز درمانی مجهز نیز احتمال بروز خطاهای تشخیصی وجود دارد[۱].

افزایش حجم تصاویر تولیدشده در سامانه‌های تصویربرداری پزشکی، رشد تعداد مراجعات بیماران و محدودیت دسترسی به متخصصان رادیولوژی، به‌ویژه در مراکز درمانی پرتردد، نیاز به استفاده از ابزارهای کمک‌تشخیصی هوشمند را بیش از گذشته آشکار کرده است. هدف این سامانه‌ها جایگزینی پزشک نیست، بلکه فراهم کردن یک نظر کمکی برای کاهش احتمال نادیده‌ماندن ضایعات، افزایش یکنواختی ارزیابی‌ها و پشتیبانی از فرآیند تصمیم‌گیری بالینی است. به همین دلیل، توسعه سامانه‌های مبتنی بر هوش مصنوعی در سال‌های اخیر به یکی از محورهای اصلی پژوهش در حوزه تصویربرداری پزشکی تبدیل شده است[۱].

پیشرفت یادگیری عمیق[۱] طی سال‌های اخیر تحول قابل‌توجهی در تحلیل خودکار تصاویر ایجاد کرده است. برخلاف روش‌های سنتی که به استخراج دستی ویژگی‌ها وابسته بودند، شبکه‌های عصبی عمیق قادرند ویژگی‌های مؤثر را مستقیماً از داده‌های خام استخراج کرده و الگوهای پیچیده موجود در تصاویر را با دقت بالایی مدل‌سازی کنند. این قابلیت موجب شده است که یادگیری عمیق در بسیاری از مسائل بینایی ماشین[۲]، از جمله طبقه‌بندی تصاویر، بخش‌بندی، آشکارسازی ناهنجاری‌ها و تحلیل تصاویر پزشکی، عملکرد موفقی از خود نشان دهد[۲].

---

[۱] Deep Learning
[۲] Computer Vision

در میان شاخه‌های مختلف بینایی ماشین، آشکارسازی اشیا[۳] جایگاه ویژه‌ای در تصویربرداری پزشکی دارد؛ زیرا علاوه بر تشخیص وجود یک ناهنجاری، محل دقیق آن را نیز مشخص می‌کند. در مسئله شکستگی‌های استخوانی، اطلاع از موقعیت دقیق شکستگی به اندازه تشخیص آن اهمیت دارد، زیرا تصمیم‌های درمانی پزشک، از جمله انتخاب روش درمان، انجام بررسی‌های تکمیلی و برنامه پیگیری بیمار، به محل و گستردگی آسیب وابسته است. ازاین‌رو، مدل‌های آشکارسازی اشیا نسبت به مدل‌های صرفاً طبقه‌بندی‌کننده، ظرفیت بیشتری برای استفاده در سامانه‌های کمک‌تشخیصی دارند[۲].

خانواده مدل‌های You Only Look Once (YOLO) یکی از شناخته‌شده‌ترین چارچوب‌های آشکارسازی اشیا محسوب می‌شود که به دلیل ایجاد تعادل میان دقت و سرعت پردازش، در کاربردهای متعددی مورد استفاده قرار گرفته است. طی سال‌های اخیر نسخه‌های مختلف این خانواده با هدف افزایش دقت، بهبود کارایی و کاهش هزینه محاسباتی توسعه یافته‌اند[۲]–[۴]. با این حال، موفقیت یک مدل در مجموعه‌داده‌های عمومی بینایی ماشین، لزوماً به معنای عملکرد مشابه آن در تصاویر پزشکی نیست. تفاوت ماهوی تصاویر پزشکی با تصاویر طبیعی از نظر ویژگی‌های بصری، ساختار داده و معیارهای ارزیابی، ضرورت بررسی مستقل این مدل‌ها در کاربردهای پزشکی را آشکار می‌سازد[۱] [۳] [۴].

بر همین اساس، پژوهش حاضر به ارزیابی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی می‌پردازد. هدف از این پژوهش، بررسی میزان کارایی این مدل در یک کاربرد واقعی پزشکی و تحلیل قابلیت استفاده از آن به‌عنوان بخشی از سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی است.

۲-۱ بیان مسئله

شکستگی‌های مچ دست یکی از شایع‌ترین آسیب‌های اسکلتی در کودکان به شمار می‌روند و سهم قابل‌توجهی از مراجعات به بخش‌های اورژانس و مراکز درمانی را به خود اختصاص می‌دهند. تشخیص صحیح این آسیب‌ها در نخستین مراجعه، نقش مهمی در انتخاب روش درمان، کاهش احتمال بروز عوارض و بازگرداندن عملکرد طبیعی اندام دارد. در عمل، رادیوگرافی نخستین ابزار تشخیصی برای ارزیابی این بیماران محسوب می‌شود و تصمیم‌گیری اولیه پزشک عمدتاً بر اساس اطلاعات حاصل از همین تصاویر انجام می‌شود[۱].

با وجود کاربرد گسترده رادیوگرافی، تفسیر تصاویر مچ دست کودکان همواره با دشواری‌هایی همراه است. برخلاف بزرگسالان، استخوان‌های کودکان در حال رشد هستند و وجود صفحات رشد، مراکز استخوان‌سازی ثانویه و تغییرات طبیعی وابسته به سن، سبب می‌شود برخی ساختارهای طبیعی ظاهری مشابه شکستگی داشته باشند یا برعکس، برخی شکستگی‌های واقعی به‌سادگی قابل تشخیص نباشند. علاوه بر این، هم‌پوشانی ساختارهای استخوانی، تفاوت کیفیت تصاویر، وضعیت قرارگیری اندام هنگام تصویربرداری و ظرافت برخی شکستگی‌ها، فرآیند تشخیص را پیچیده‌تر می‌کنند[۱].

[۳] Object Detection

در نتیجه، تشخیص این آسیب‌ها تنها به کیفیت تصاویر وابسته نیست، بلکه به تجربه، دانش تخصصی و قضاوت بالینی پزشک نیز بستگی دارد. در مراکز درمانی پرتردد یا مراکزی که دسترسی محدودتری به متخصصان رادیولوژی کودکان دارند، احتمال نادیده‌ماندن شکستگی‌های ظریف یا اختلاف نظر میان ارزیابان افزایش می‌یابد. چنین شرایطی می‌تواند موجب انجام تصویربرداری‌های تکمیلی، تأخیر در آغاز درمان، انتخاب روش درمانی نامناسب و در برخی موارد بروز عوارض ناشی از تشخیص دیرهنگام شود[۱].

در پاسخ به این چالش‌ها، استفاده از سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی طی سال‌های اخیر به یکی از رویکردهای مهم در تحلیل تصاویر پزشکی تبدیل شده است. این سامانه‌ها با تحلیل خودکار تصاویر، قادرند نواحی مشکوک را شناسایی کرده و به‌عنوان ابزاری کمکی در کنار پزشک مورد استفاده قرار گیرند. در این میان، آشکارسازی اشیا به دلیل توانایی هم‌زمان در تشخیص و تعیین موقعیت مکانی ضایعات، برای کاربردهایی مانند تشخیص شکستگی اهمیت ویژه‌ای دارد؛ زیرا در این مسئله، تعیین محل دقیق شکستگی به اندازه تشخیص وجود آن اهمیت دارد[۲].

پیشرفت معماری‌های آشکارسازی اشیا موجب معرفی نسل‌های جدیدی از مدل‌های خانواده یولو شده است که هر یک با هدف بهبود دقت، سرعت و کارایی محاسباتی توسعه یافته‌اند[۲]–[۴]. با این حال، عملکرد مطلوب این مدل‌ها در مجموعه‌داده‌های عمومی بینایی ماشین، به‌تنهایی نمی‌تواند نشان‌دهنده کارایی آن‌ها در تصاویر پزشکی باشد. تفاوت‌های اساسی میان تصاویر طبیعی و تصاویر رادیوگرافی، از جمله نوع اطلاعات، ساختار تصویر، ویژگی‌های بافتی و معیارهای تشخیص، ایجاب می‌کند که این مدل‌ها در مجموعه‌داده‌های تخصصی پزشکی نیز به‌صورت مستقل ارزیابی شوند[۱] [۳] [۴].

در سال‌های اخیر، مجموعه‌داده بیمارستان گراس به یکی از منابع مرجع برای آموزش و ارزیابی مدل‌های یادگیری عمیق در زمینه شکستگی‌های مچ دست کودکان تبدیل شده است[۱]. در کنار آن، مطالعات مختلفی عملکرد نسخه‌های گوناگون خانواده یولو را در این مسئله بررسی کرده‌اند[۲]–[۴]. با وجود این، معرفی نسخه‌های جدید این خانواده، نیاز به ارزیابی مستمر آن‌ها بر روی مجموعه‌داده‌های استاندارد و با استفاده از معیارهای پذیرفته‌شده را همچنان حفظ کرده است؛ زیرا تنها از طریق چنین ارزیابی‌هایی می‌توان درباره قابلیت استفاده از این مدل‌ها در کاربردهای واقعی پزشکی قضاوت کرد.

بر همین اساس، مسئله اصلی این پژوهش، ارزیابی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی است. این پژوهش با بهره‌گیری از یک مجموعه‌داده استاندارد و معیارهای متداول ارزیابی، تلاش می‌کند میزان کارایی، قابلیت‌ها و محدودیت‌های این مدل را در یک کاربرد واقعی پزشکی بررسی کرده و شواهد تجربی لازم را برای تحلیل امکان استفاده از آن در سامانه‌های کمک‌تشخیصی فراهم آورد.

۳-۱ اهمیت پژوهش

تشخیص دقیق شکستگی‌های مچ دست در کودکان تنها به شناسایی وجود یا عدم وجود یک آسیب محدود نمی‌شود، بلکه مبنای بسیاری از تصمیم‌های درمانی بعدی است. انتخاب روش درمان، مدت زمان بی‌حرکت‌سازی اندام، نیاز به انجام تصویربرداری‌های تکمیلی، ارجاع بیمار به جراح ارتوپد و برنامه پیگیری درمان، همگی به صحت تشخیص اولیه وابسته هستند. ازاین‌رو، هرگونه بهبود در فرآیند تشخیص می‌تواند علاوه بر افزایش کیفیت خدمات درمانی، احتمال بروز خطاهای تشخیصی، تأخیر در درمان و عوارض ناشی از آن را کاهش دهد. به همین دلیل، توسعه ابزارهایی که بتوانند در کنار پزشک به تفسیر دقیق‌تر تصاویر رادیوگرافی کمک کنند، از دیدگاه بالینی دارای اهمیت فراوان است[۱].

با وجود پیشرفت چشمگیر روش‌های مبتنی بر یادگیری عمیق، انتقال موفق این فناوری از مجموعه‌داده‌های عمومی بینایی ماشین به حوزه تصویربرداری پزشکی، همواره با چالش‌هایی همراه بوده است. تصاویر رادیوگرافی از نظر ویژگی‌های بصری، نحوه تشکیل تصویر، ساختار بافت، ابعاد نواحی مورد بررسی و معیارهای تشخیصی، تفاوت‌های اساسی با تصاویر طبیعی دارند. در نتیجه، نمی‌توان انتظار داشت مدلی که در مجموعه‌داده‌های عمومی عملکرد مطلوبی دارد، بدون ارزیابی مستقل در تصاویر پزشکی نیز همان عملکرد را ارائه دهد. از این منظر، بررسی قابلیت تعمیم[۴] مدل‌های آشکارسازی اشیا به تصاویر پزشکی، یکی از موضوعات مهم پژوهش در حوزه هوش مصنوعی پزشکی محسوب می‌شود[۲] [۳].

اهمیت پژوهشی این مطالعه نیز از همین موضوع ناشی می‌شود. توسعه مستمر معماری‌های جدید در خانواده مدل‌های یولو، فرصت مناسبی برای افزایش دقت و کارایی سامانه‌های بینایی ماشین فراهم کرده است، اما ارزش واقعی هر معماری جدید تنها زمانی مشخص می‌شود که در کاربردهای تخصصی و بر اساس مجموعه‌داده‌ها و معیارهای ارزیابی استاندارد مورد بررسی قرار گیرد. ارزیابی تجربی مدل‌های جدید بر روی داده‌های پزشکی، علاوه بر فراهم کردن شواهد قابل استناد برای جامعه علمی، امکان مقایسه عملکرد معماری‌های مختلف را در شرایط یکسان فراهم می‌کند و به درک بهتر قابلیت‌ها و محدودیت‌های آن‌ها کمک می‌نماید[۲].

از دیدگاه کاربردی، پژوهش حاضر به دنبال ارائه یک سامانه تشخیصی آماده استفاده در محیط‌های درمانی نیست، بلکه هدف آن بررسی ظرفیت یک معماری نوین برای استفاده در چنین سامانه‌هایی است. نتایج حاصل از این پژوهش می‌تواند به‌عنوان یکی از شواهد موردنیاز در فرآیند توسعه سامانه‌های کمک‌تشخیصی[۵] یا CAD مورد استفاده قرار گیرد. در صورت دستیابی به عملکرد مناسب، چنین مدل‌هایی می‌توانند در آینده به‌عنوان ابزاری کمکی، نواحی مشکوک به شکستگی را پیش از بررسی نهایی پزشک مشخص کنند و از این طریق، به افزایش یکنواختی ارزیابی‌ها، کاهش احتمال نادیده‌ماندن ضایعات و بهبود فرآیند تصمیم‌گیری بالینی کمک نمایند. با این حال، تصمیم‌گیری نهایی همواره بر عهده پزشک خواهد بود و نقش این سامانه‌ها، پشتیبانی از فرآیند تشخیص است، نه جایگزینی آن[۵] .

---

[۴] Generalization

[۵] Computer-Aided Diagnosis

در مجموع، اهمیت این پژوهش را می‌توان در سه بعد اصلی خلاصه کرد: بعد بالینی که بر بهبود فرآیند تشخیص و حمایت از تصمیم‌گیری پزشک تمرکز دارد؛ بعد پژوهشی که به ارزیابی عملکرد یک معماری نوین در یک مسئله تخصصی پزشکی می‌پردازد؛ و بعد کاربردی که زمینه استفاده از یافته‌های پژوهش در توسعه سامانه‌های کمک‌تشخیصی آینده را فراهم می‌کند. از این رو، پژوهش حاضر علاوه بر پرداختن به یک مسئله مهم در تصویربرداری پزشکی، می‌تواند در توسعه دانش مرتبط با کاربرد مدل‌های آشکارسازی اشیا در حوزه پزشکی نیز نقش مؤثری ایفا کند.

۱-۴ اهداف پژوهش

با توجه به چالش‌های موجود در تشخیص شکستگی‌های مچ دست کودکان، اهمیت توسعه ابزارهای کمک‌تشخیصی مبتنی بر هوش مصنوعی و ضرورت ارزیابی مستقل مدل‌های نوین آشکارسازی اشیا در کاربردهای پزشکی، این پژوهش با هدف بررسی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی انجام شده است. این پژوهش تلاش می‌کند با استفاده از یک مجموعه‌داده استاندارد و معیارهای پذیرفته‌شده ارزیابی، تصویری روشن از توانایی‌ها و محدودیت‌های این مدل در یک کاربرد واقعی پزشکی ارائه دهد. بر این اساس، هدف اصلی و اهداف اختصاصی پژوهش به شرح زیر تعریف می‌شوند.

هدف اصلی

ارزیابی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی.

اهداف اختصاصی

- ارزیابی توانایی مدل در شناسایی خودکار شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی.
- سنجش عملکرد مدل با استفاده از معیارهای استاندارد ارزیابی
- تحلیل نقاط قوت و محدودیت‌های مدل در تشخیص و مکان‌یابی شکستگی‌ها.
- بررسی قابلیت استفاده از مدل در توسعه سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی برای تحلیل تصاویر رادیوگرافی.
- ارائه شواهد تجربی برای ارزیابی قابلیت به‌کارگیری معماری‌های جدید آشکارسازی اشیا در کاربردهای تخصصی تصویربرداری پزشکی.

اهداف فوق به گونه‌ای انتخاب شده‌اند که تمامی آن‌ها در ادامه پژوهش و بر اساس نتایج حاصل از ارزیابی مدل قابل بررسی و تحلیل باشند و در مجموع، تصویری جامع از عملکرد مدل منتخب در این کاربرد تخصصی ارائه دهند.

۱-۵ سؤالات پژوهش

با توجه به بیان مسئله، اهداف پژوهش و ضرورت ارزیابی مستقل مدل‌های نوین آشکارسازی اشیا در کاربردهای پزشکی، این پژوهش در پی پاسخ‌گویی به مجموعه‌ای از پرسش‌های مشخص و قابل ارزیابی است. این پرسش‌ها به گونه‌ای تدوین شده‌اند که پاسخ به آن‌ها از طریق نتایج حاصل از آموزش، ارزیابی و تحلیل عملکرد مدل امکان‌پذیر باشد و در مجموع، تصویری جامع از قابلیت‌ها، محدودیت‌ها و ظرفیت استفاده از مدل در این کاربرد تخصصی ارائه دهد. بر این اساس، سؤالات اصلی پژوهش عبارت‌اند از:

- سؤال اول: عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی چگونه است؟
- سؤال دوم: عملکرد مدل از نظر معیارهای استاندارد ارزیابی آشکارسازی اشیا، از جمله دقت مثبت، فراخوانی و میانگین دقت (mAP)در چه سطحی قرار دارد؟
- سؤال سوم: مهم‌ترین نقاط قوت و محدودیت‌های مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان چیست؟
- سؤال چهارم: تا چه اندازه می‌توان از نتایج این پژوهش به‌عنوان مبنایی برای توسعه سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی در تحلیل تصاویر رادیوگرافی استفاده کرد؟

پاسخ به این پرسش‌ها علاوه بر ارزیابی کمی عملکرد مدل، امکان تحلیل قابلیت‌های آن در یک کاربرد واقعی پزشکی را فراهم می‌سازد و می‌تواند مسیر مناسبی برای انجام پژوهش‌های تکمیلی در این حوزه ترسیم کند.

۱-۶ دستاوردهای پژوهش

پژوهش حاضر از نوع پژوهش کاربردی است و هدف آن طراحی یک معماری جدید یا توسعه یک الگوریتم نوین نیست، بلکه تمرکز آن بر ارزیابی نظام‌مند عملکرد یک معماری نوین آشکارسازی اشیا در یک مسئله تخصصی تصویربرداری پزشکی است. بنابراین، ارزش علمی این پژوهش را باید در دستاوردهای حاصل از ارزیابی، تحلیل و مستندسازی عملکرد مدل جست‌وجو کرد. مهم‌ترین دستاوردهای این پژوهش عبارت‌اند از:

۱ -ارزیابی عملکرد مدل یولو ۲۶ در یک کاربرد تخصصی پزشکی: در این پژوهش، عملکرد مدل یولو ۲۶ در مسئله تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان مورد بررسی قرار گرفته است. این ارزیابی تصویری روشن از توانایی مدل در تحلیل تصاویر رادیوگرافی و میزان قابلیت استفاده از آن در این کاربرد پزشکی ارائه می‌دهد.

۲ -انجام ارزیابی بر پایه یک مجموعه‌داده استاندارد پزشکی: استفاده از مجموعه‌داده‌های معتبر در حوزه شکستگی‌های مچ دست کودکان، امکان انجام یک ارزیابی ساختاریافته و مبتنی بر معیارهای پذیرفته‌شده را فراهم کرده است. این موضوع قابلیت مقایسه نتایج پژوهش با سایر مطالعات منتشرشده در این حوزه را نیز افزایش می‌دهد.

۳ -تحلیل جامع عملکرد مدل: در این پژوهش، علاوه بر ارزیابی کمی مدل، عملکرد آن از جنبه‌های مختلف شامل توانایی تشخیص، کیفیت مکان‌یابی، نقاط قوت و محدودیت‌های مشاهده‌شده نیز مورد تحلیل قرار گرفته است. این تحلیل دید دقیق‌تری نسبت به قابلیت‌های مدل در تصاویر رادیوگرافی کودکان فراهم می‌آورد و می‌تواند در انتخاب معماری‌های مناسب برای پژوهش‌های آینده مفید باشد.

۴ -فراهم کردن شواهد تجربی برای پژوهش‌های آینده: یافته‌های این پژوهش می‌تواند به‌عنوان یک مرجع تجربی برای مطالعات آینده در زمینه کاربرد مدل‌های آشکارسازی اشیا در تحلیل تصاویر پزشکی مورد استفاده قرار گیرد. همچنین نتایج حاصل از این پژوهش می‌تواند در توسعه و ارزیابی سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی و انتخاب معماری‌های مناسب برای کاربردهای مشابه نقش مؤثری ایفا کند.

در مجموع، دستاورد اصلی این پژوهش ارائه یک ارزیابی تجربی، نظام‌مند و قابل استناد از عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان است. این پژوهش ضمن تحلیل قابلیت‌ها و محدودیت‌های مدل در یک مسئله واقعی پزشکی، شواهد لازم برای ادامه پژوهش‌ها در زمینه به‌کارگیری مدل‌های آشکارسازی اشیا در تصویربرداری پزشکی را فراهم می‌کند.

۱-۷ ساختار گزارش

این گزارش در شش فصل تدوین شده است که هر فصل به یکی از جنبه‌های اصلی پژوهش اختصاص دارد و در مجموع، روند انجام پژوهش را از معرفی مسئله تا تحلیل نتایج و جمع‌بندی نهایی تشریح می‌کند.

- فصل اول با عنوان کلیات پژوهش، به معرفی موضوع، بیان مسئله، اهمیت پژوهش، اهداف، سؤالات، دستاوردهای پژوهش و ساختار کلی گزارش اختصاص دارد و چارچوب نظری و منطقی انجام پژوهش را تبیین می‌کند.
- فصل دوم با عنوان مبانی نظری و مرور پژوهش‌ها، مفاهیم پایه موردنیاز پژوهش شامل تصویربرداری رادیوگرافی، شکستگی‌های مچ دست کودکان، یادگیری عمیق، آشکارسازی اشیا، خانواده مدل‌های یولو و مجموعه‌داده GRAZPEDWRI-DX را معرفی کرده و پژوهش‌های پیشین مرتبط را مرور و تحلیل می‌کند.
- فصل سوم با عنوان روش پژوهش، روش انجام پژوهش، نحوه آماده‌سازی داده‌ها، محیط سخت‌افزاری و نرم‌افزاری، تنظیمات آموزش مدل، معیارهای ارزیابی و مراحل اجرای آزمایش‌ها را تشریح می‌کند.
- فصل چهارم با عنوان پیاده‌سازی و اجرای مدل، فرآیند عملی اجرای پژوهش، آموزش مدل، اعتبارسنجی، مدیریت داده‌ها و جزئیات پیاده‌سازی را ارائه می‌دهد.
- فصل پنجم با عنوان نتایج و تحلیل، نتایج حاصل از ارزیابی مدل را بر اساس معیارهای استاندارد ارائه کرده و عملکرد آن را از جنبه‌های مختلف مورد تحلیل و بررسی قرار می‌دهد. همچنین نقاط قوت، محدودیت‌ها و یافته‌های حاصل از پژوهش در این فصل بحث می‌شوند.

- در نهایت، فصل ششم با عنوان جمع‌بندی و پیشنهادها، مهم‌ترین یافته‌های پژوهش را جمع‌بندی کرده، میزان تحقق اهداف و پاسخ به سؤالات پژوهش را بررسی می‌کند و پیشنهادهایی برای ادامه پژوهش و توسعه مطالعات آینده ارائه می‌دهد.

فصل دوم مبانی نظری و مرور پژوهش‌ها

۲-۱ مقدمه

در فصل نخست، مسئله پژوهش، اهمیت آن، اهداف، سؤالات و دستاوردهای مورد انتظار تبیین شد و ضرورت ارزیابی مدل‌های نوین آشکارسازی اشیا در تشخیص شکستگی‌های مچ دست کودکان مورد بررسی قرار گرفت. تحقق اهداف پژوهش و تحلیل صحیح نتایج حاصل از آن، مستلزم شناخت مبانی نظری مرتبط با مسئله و بررسی وضعیت پژوهش‌های انجام‌شده در این حوزه است. ازاین‌رو، پیش از تشریح روش انجام پژوهش، لازم است مفاهیم پایه، فناوری‌های مورد استفاده و مطالعات پیشین به‌صورت نظام‌مند بررسی شوند.

پژوهش حاضر در تقاطع چند حوزه تخصصی شامل تصویربرداری پزشکی، تشخیص شکستگی‌های مچ دست کودکان، یادگیری عمیق، آشکارسازی اشیا و سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی قرار دارد. درک صحیح ارتباط میان این حوزه‌ها، زمینه لازم برای تحلیل روش پژوهش، تفسیر نتایج و ارزیابی عملکرد مدل را فراهم می‌کند. به همین دلیل، در بخش نخست این فصل، مبانی نظری موردنیاز برای انجام پژوهش معرفی می‌شوند و مفاهیمی بررسی خواهند شد که در ادامه پژوهش به‌طور مستقیم مورد استفاده قرار می‌گیرند.

علاوه بر مبانی نظری، بررسی پژوهش‌های پیشین نیز نقش مهمی در شناخت وضعیت موجود، دستاوردهای مطالعات گذشته و چالش‌های باقی‌مانده دارد. تحلیل مطالعات منتشرشده نشان می‌دهد که با وجود پیشرفت قابل‌توجه مدل‌های مبتنی بر یادگیری عمیق در تحلیل تصاویر پزشکی، همچنان ارزیابی مستقل معماری‌های جدید بر روی مجموعه‌داده‌های استاندارد پزشکی از اهمیت ویژه‌ای برخوردار است. بررسی پژوهش‌های پیشین علاوه بر فراهم کردن امکان مقایسه پژوهش حاضر با مطالعات مشابه، مبنایی برای تبیین جایگاه علمی این پژوهش و تحلیل نتایج حاصل از آن خواهد بود.

بر این اساس، این فصل در دو بخش اصلی سازمان‌دهی شده است. در بخش نخست، مبانی نظری مرتبط با موضوع پژوهش، شامل تصویربرداری رادیوگرافی، شکستگی‌های مچ دست کودکان، یادگیری عمیق، آشکارسازی اشیا، خانواده مدل‌های یولو، مجموعه‌داده و معیارهای ارزیابی معرفی می‌شوند. در بخش دوم، پژوهش‌های پیشین مرتبط با کاربرد مدل‌های یادگیری عمیق در تشخیص شکستگی‌های استخوانی و به‌ویژه شکستگی‌های مچ دست کودکان به‌صورت تحلیلی بررسی خواهند شد. در پایان فصل نیز، با جمع‌بندی مبانی نظری و تحلیل مطالعات پیشین، جایگاه پژوهش حاضر و خلأ پژوهشی موجود تبیین خواهد شد.

۲-۲ شکستگی‌های مچ دست در کودکان

شکستگی‌های مچ دست از شایع‌ترین آسیب‌های اسکلتی در دوران کودکی به شمار می‌روند و بخش قابل‌توجهی از مراجعات کودکان به مراکز درمانی و بخش‌های اورژانس را تشکیل می‌دهند. بروز این آسیب‌ها عمدتاً با زمین‌خوردن هنگام بازی، فعالیت‌های ورزشی یا وارد شدن ضربه مستقیم به اندام فوقانی مرتبط است. از آنجا که استخوان‌های کودکان هنوز در حال رشد هستند، تشخیص صحیح و به‌موقع این آسیب‌ها اهمیت ویژه‌ای دارد؛ زیرا تصمیم‌گیری درباره نوع درمان، مدت بی‌حرکت‌سازی، نیاز به بررسی‌های تکمیلی و برنامه پیگیری بیمار بر پایه نتایج حاصل از ارزیابی اولیه انجام می‌شود[۱].

از دیدگاه آناتومیکی، مچ دست از بخش انتهایی استخوان‌های رادیوس[۶] و اولنا[۷] و مجموعه‌ای از استخوان‌های کارپ تشکیل شده است. تفاوت اصلی مچ دست کودکان با بزرگسالان، وجود صفحات رشد و ادامه فرایند استخوان‌سازی است. این ساختارها در تصاویر رادیوگرافی ظاهری متفاوت از استخوان بالغ دارند و بخشی طبیعی از روند رشد محسوب می‌شوند. در نتیجه، ظاهر طبیعی استخوان در کودکان می‌تواند با برخی علائم شکستگی شباهت داشته باشد یا برعکس، برخی شکستگی‌های واقعی را پنهان کند[۱].

به همین دلیل، چالش اصلی در ارزیابی تصاویر رادیوگرافی کودکان، صرفاً تشخیص وجود شکستگی نیست، بلکه تمایز میان ساختارهای طبیعی در حال رشد و ضایعات واقعی استخوانی است. وجود صفحات رشد، مراکز استخوان‌سازی، هم‌پوشانی ساختارهای استخوانی، کیفیت متفاوت تصاویر و ظرافت برخی شکستگی‌ها موجب می‌شود که حتی در تصاویر استاندارد نیز تشخیص صحیح همواره آسان نباشد. علاوه بر این، برخی شکستگی‌ها تنها با تغییرات بسیار جزئی در پیوستگی قشر استخوان یا تغییر شکل محدود استخوان همراه هستند که تشخیص آن‌ها نیازمند دقت و تجربه بالینی است [۱]،[۲].

پیچیدگی‌های یادشده سبب می‌شود که فرآیند تشخیص تا حد زیادی به تجربه متخصص وابسته باشد و در برخی موارد میان ارزیابان مختلف اختلاف نظر وجود داشته باشد. از سوی دیگر، عدم تشخیص به‌موقع شکستگی یا آسیب صفحات رشد می‌تواند به انتخاب درمان نامناسب، تأخیر در بهبود، تغییر شکل استخوان یا اختلال در رشد طبیعی اندام منجر شود. در مقابل، تشخیص نادرست شکستگی نیز ممکن است باعث انجام اقدامات درمانی غیرضروری و افزایش هزینه‌های درمان شود. بنابراین، افزایش دقت و یکنواختی فرآیند تشخیص، یکی از نیازهای مهم در مدیریت بیماران کودک مبتلا به آسیب‌های مچ دست محسوب می‌شود[۱].

با توجه به این چالش‌ها، تصویربرداری رادیوگرافی همچنان نخستین و رایج‌ترین روش ارزیابی شکستگی‌های مچ دست در کودکان است. با این حال، همان ویژگی‌هایی که موجب ارزش بالینی این روش شده‌اند، فرآیند تفسیر تصاویر را نیز پیچیده کرده‌اند. ازاین‌رو، شناخت قابلیت‌ها و محدودیت‌های رادیوگرافی و ویژگی‌های تصاویر حاصل از آن، پیش‌نیاز بررسی روش‌های مبتنی بر هوش مصنوعی

---

[۶] **Radius**
[۷] **Ulna**

برای تحلیل این تصاویر به شمار می‌رود. به همین دلیل، در بخش بعدی، اصول تصویربرداری رادیوگرافی و ویژگی‌های آن در تشخیص شکستگی‌های مچ دست کودکان بررسی خواهد شد.

۳-۲ تصویربرداری رادیوگرافی در تشخیص شکستگی‌های مچ دست

رادیوگرافی رایج‌ترین و نخستین روش تصویربرداری برای ارزیابی آسیب‌های اسکلتی محسوب می‌شود و در اغلب مراکز درمانی به‌عنوان ابزار استاندارد اولیه برای بررسی شکستگی‌های استخوانی مورد استفاده قرار می‌گیرد. این روش بر پایه عبور پرتوهای ایکس از بافت‌های بدن و ثبت میزان جذب آن‌ها توسط ساختارهای مختلف عمل می‌کند. تفاوت در میزان جذب پرتو توسط استخوان و بافت‌های نرم، امکان مشاهده ساختار استخوان‌ها و تشخیص بسیاری از آسیب‌های اسکلتی را فراهم می‌سازد. به دلیل سرعت بالا، دسترس‌پذیری گسترده، هزینه نسبتاً پایین و امکان انجام سریع در شرایط اورژانسی، رادیوگرافی همچنان اولین انتخاب پزشکان برای ارزیابی شکستگی‌های مچ دست در کودکان است[۱][۲].

با وجود کاربرد گسترده، رادیوگرافی دارای محدودیت‌هایی است که بر فرآیند تشخیص تأثیر می‌گذارند. تصاویر رادیوگرافی، نمایش دوبعدی از ساختارهای سه‌بعدی بدن هستند؛ ازاین‌رو، هم‌پوشانی استخوان‌ها و سایر ساختارهای آناتومیکی می‌تواند مشاهده برخی شکستگی‌ها را دشوار کند. علاوه بر این، کیفیت تصاویر تحت تأثیر عواملی مانند نحوه قرارگیری اندام، شرایط تصویربرداری، میزان تابش پرتو و حرکت بیمار قرار می‌گیرد. این عوامل موجب می‌شوند کیفیت اطلاعات موجود در تصاویر مختلف یکسان نباشد و فرآیند تفسیر آن‌ها با چالش همراه شود[۲].

این چالش‌ها در تصاویر رادیوگرافی کودکان پیچیده‌تر نیز می‌شوند. وجود صفحات رشد، مراکز استخوان‌سازی فعال و تغییرات طبیعی وابسته به سن سبب می‌شود که ظاهر استخوان‌های کودک با بزرگسالان تفاوت قابل‌توجهی داشته باشد. در نتیجه، برخی ساختارهای طبیعی ممکن است در تصاویر رادیوگرافی مشابه شکستگی به نظر برسند و در مقابل، برخی شکستگی‌های واقعی تنها با تغییرات بسیار ظریف در پیوستگی استخوان یا تغییر شکل محدود قشر استخوان همراه باشند. بنابراین، تشخیص صحیح شکستگی در کودکان علاوه بر کیفیت تصویر، به شناخت دقیق ویژگی‌های آناتومیکی و تجربه متخصص نیز وابسته است[۱] [۳].

از دیدگاه پردازش تصویر، تصاویر رادیوگرافی ویژگی‌هایی دارند که آن‌ها را از بسیاری از تصاویر طبیعی متمایز می‌کند. کنتراست محدود برخی نواحی، شباهت شدت روشنایی بافت‌های مختلف، اندازه کوچک ضایعات، تنوع شرایط تصویربرداری و تفاوت‌های ناشی از سن بیمار، استخراج ویژگی‌های مؤثر را با دشواری همراه می‌سازد. به همین دلیل، روش‌هایی که برای تحلیل تصاویر طبیعی توسعه یافته‌اند، لزوماً بدون ارزیابی و سازگارسازی در تصاویر پزشکی عملکرد مشابهی نخواهند داشت [۳] [۴].

با وجود این محدودیت‌ها، رادیوگرافی همچنان مرجع اصلی تشخیص اولیه شکستگی‌های مچ دست محسوب می‌شود و تقریباً تمامی مطالعات مرتبط با توسعه سامانه‌های کمک‌تشخیصی در این حوزه بر پایه تصاویر رادیوگرافی انجام شده‌اند. ازاین‌رو، هرگونه روش

هوشمند برای تحلیل خودکار این تصاویر باید بتواند ضمن حفظ دقت تشخیص، با چالش‌هایی نظیر هم‌پوشانی ساختارها، تغییرات طبیعی استخوان‌های در حال رشد و کیفیت متغیر تصاویر نیز سازگار باشد. این موضوع، استفاده از روش‌های پیشرفته یادگیری عمیق را به یکی از رویکردهای اصلی پژوهش‌های اخیر در تحلیل تصاویر پزشکی تبدیل کرده است [۴] [۵].

بر این اساس، داده مورد استفاده در پژوهش حاضر نیز مجموعه‌ای از تصاویر رادیوگرافی استاندارد مچ دست کودکان است که هدف از تحلیل آن‌ها، تشخیص و مکان‌یابی دقیق نواحی شکستگی است. دستیابی به این هدف مستلزم استفاده از روش‌هایی است که بتوانند ویژگی‌های پیچیده و الگوهای ظریف موجود در تصاویر را به‌صورت خودکار استخراج و تحلیل کنند. در سال‌های اخیر، یادگیری عمیق با توانایی بالا در استخراج سلسله‌مراتبی ویژگی‌های تصویری، به یکی از مؤثرترین رویکردها برای حل چنین مسائلی تبدیل شده است. ازاین‌رو، در بخش بعدی، مبانی یادگیری عمیق و کاربرد آن در تحلیل تصاویر پزشکی مورد بررسی قرار می‌گیرد.

۲-۴ یادگیری عمیق و کاربرد آن در تحلیل تصاویر پزشکی

پیشرفت فناوری‌های پردازش تصویر و افزایش توان محاسباتی سامانه‌های رایانه‌ای، زمینه را برای استفاده گسترده از هوش مصنوعی در تحلیل تصاویر پزشکی فراهم کرده است. در سال‌های اخیر، هوش مصنوعی به یکی از مهم‌ترین فناوری‌های تحول‌آفرین در حوزه سلامت تبدیل شده و در کاربردهایی نظیر تشخیص بیماری، پیش‌بینی پیامدهای درمان، پردازش سیگنال‌های زیستی و تحلیل تصاویر پزشکی مورد استفاده قرار گرفته است. هدف از به‌کارگیری این فناوری‌ها، جایگزینی پزشک نیست، بلکه افزایش دقت، سرعت و یکنواختی فرآیندهای تشخیصی و فراهم کردن ابزارهای تصمیم‌یار برای متخصصان است [۶] [۷].

یکی از شاخه‌های اصلی هوش مصنوعی، یادگیری ماشین است که در آن مدل‌ها به‌جای پیروی از مجموعه‌ای از قوانین از پیش تعریف‌شده، الگوهای موجود در داده‌ها را فرا می‌گیرند. در روش‌های کلاسیک یادگیری ماشین، استخراج ویژگی‌های مناسب از تصاویر معمولاً به‌صورت دستی یا با استفاده از الگوریتم‌های مهندسی‌شده انجام می‌شود. کیفیت این ویژگی‌ها تأثیر مستقیمی بر عملکرد مدل دارد و در بسیاری از مسائل پیچیده، طراحی ویژگی‌های مناسب به دانش تخصصی و تجربه فراوان نیازمند است[٦].

به‌منظور غلبه بر این محدودیت‌ها، یادگیری عمیق معرفی شد. در این رویکرد، شبکه‌های عصبی عمیق قادرند ویژگی‌های موردنیاز را به‌صورت خودکار و به‌صورت سلسله‌مراتبی از داده‌های خام استخراج کنند. برخلاف روش‌های کلاسیک که کیفیت مدل تا حد زیادی به کیفیت ویژگی‌های طراحی‌شده وابسته است، در یادگیری عمیق فرآیند استخراج ویژگی و یادگیری مدل به‌صورت هم‌زمان انجام می‌شود. این ویژگی موجب شده است که یادگیری عمیق در بسیاری از مسائل بینایی ماشین، به‌ویژه تحلیل تصاویر، عملکردی به‌مراتب بهتر از روش‌های سنتی ارائه دهد [۷] [۸].

در حوزه تحلیل تصاویر، شبکه‌های عصبی کانولوشنی[۸] یا CNN به‌عنوان یکی از مهم‌ترین معماری‌های یادگیری عمیق شناخته می‌شوند. این شبکه‌ها با استفاده از عملگرهای کانولوشن، قادرند الگوهای محلی موجود در تصویر را استخراج کرده و در لایه‌های عمیق‌تر، آن‌ها را به ویژگی‌های پیچیده‌تر تبدیل کنند. چنین ساختاری امکان شناسایی خودکار لبه‌ها، بافت‌ها، اشکال و الگوهای پیچیده تصویری را فراهم می‌کند و نیاز به استخراج دستی ویژگی‌ها را از میان برمی‌دارد. به همین دلیل، CNNها به پایه اصلی بسیاری از مدل‌های مدرن تحلیل تصویر تبدیل شده‌اند[۸] .

در سال‌های اخیر، یادگیری عمیق در حوزه تصویربرداری پزشکی نیز به‌طور گسترده مورد استفاده قرار گرفته است. کاربردهای این فناوری شامل طبقه‌بندی تصاویر پزشکی، بخش‌بندی ساختارهای آناتومیکی، تشخیص ناهنجاری‌ها، آشکارسازی ضایعات و کمک به تشخیص بیماری‌ها است. نتایج مطالعات متعدد نشان داده است که در بسیاری از کاربردها، مدل‌های مبتنی بر یادگیری عمیق قادرند عملکردی قابل‌رقابت با متخصصان ارائه دهند؛ با این حال، عملکرد این مدل‌ها به عواملی مانند کیفیت داده، حجم مجموعه‌داده، نحوه برچسب‌گذاری و میزان شباهت داده‌های آموزشی و داده‌های واقعی وابسته است. ازاین‌رو، ارزیابی مستقل این مدل‌ها در هر کاربرد پزشکی ضروری است و نتایج یک مطالعه را نمی‌توان بدون بررسی به سایر حوزه‌ها تعمیم داد [۷][۹] .

با وجود موفقیت‌های چشمگیر، استفاده از یادگیری عمیق در پزشکی با چالش‌هایی نیز همراه است. نیاز به مجموعه‌داده‌های بزرگ و باکیفیت، هزینه بالای برچسب‌گذاری تصاویر توسط متخصصان، دشواری تفسیر تصمیم‌های مدل[۹]، احتمال کاهش عملکرد در داده‌های خارج از توزیع آموزشی[۱۰] و ملاحظات مربوط به قابلیت تعمیم مدل، از جمله مهم‌ترین محدودیت‌های این فناوری به شمار می‌روند. به همین دلیل، ارزیابی دقیق مدل‌ها بر روی مجموعه‌داده‌های استاندارد و تحلیل نقاط قوت و محدودیت‌های آن‌ها، بخش جدایی‌ناپذیر پژوهش‌های مبتنی بر یادگیری عمیق در پزشکی محسوب می‌شود [۸] [۹].

در میان کاربردهای مختلف یادگیری عمیق در تحلیل تصاویر پزشکی، روش‌های مبتنی بر آشکارسازی اشیا جایگاه ویژه‌ای دارند؛ زیرا این روش‌ها علاوه بر تشخیص وجود یک ناهنجاری، قادرند موقعیت مکانی آن را نیز مشخص کنند. در مسئله تشخیص شکستگی‌های مچ دست، اطلاع از محل دقیق شکستگی به اندازه تشخیص وجود آن اهمیت دارد و همین موضوع سبب شده است که استفاده از مدل‌های آشکارسازی اشیا نسبت به مدل‌های صرفاً طبقه‌بندی‌کننده، گزینه مناسب‌تری برای این پژوهش باشد. ازاین‌رو، در بخش بعدی، مبانی نظری روش‌های آشکارسازی اشیا و جایگاه آن‌ها در تحلیل تصاویر پزشکی بررسی خواهد شد.

---

[۸] **Convolutional Neural Networks**
[۹] Explainability
[۱۰] Domain Shift

۲-۵ آشکارسازی اشیا در تحلیل تصاویر پزشکی

در بسیاری از کاربردهای بینایی ماشین، هدف تنها تشخیص وجود یک شیء یا ناهنجاری در تصویر نیست، بلکه تعیین موقعیت دقیق آن نیز اهمیت دارد. به همین دلیل، بسته به نوع مسئله، وظایف مختلفی برای تحلیل تصاویر تعریف شده‌اند که هر یک خروجی متفاوتی تولید می‌کنند. انتخاب صحیح این وظیفه، یکی از مهم‌ترین مراحل طراحی سامانه‌های مبتنی بر بینایی ماشین محسوب می‌شود؛ زیرا نوع خروجی مورد انتظار، معماری مدل و معیارهای ارزیابی را تحت تأثیر قرار می‌دهد[۱۰] .

در ساده‌ترین حالت، طبقه‌بندی تصویر[۱۱] تنها وجود یا عدم وجود یک کلاس مشخص را در کل تصویر تعیین می‌کند. برای مثال، در مسئله تشخیص شکستگی، یک مدل طبقه‌بندی تنها مشخص می‌کند که آیا تصویر شامل شکستگی هست یا خیر. با وجود سادگی این رویکرد، خروجی آن برای بسیاری از کاربردهای پزشکی کافی نیست؛ زیرا اطلاعاتی درباره محل دقیق ضایعه در اختیار پزشک قرار نمی‌دهد[۱۱] .

در مقابل، آشکارسازی اشیا علاوه بر تعیین نوع شیء، موقعیت مکانی آن را نیز در تصویر مشخص می‌کند. خروجی این روش معمولاً شامل یک یا چند جعبه محدودکننده[۱۲] به همراه برچسب کلاس و میزان اطمینان مدل است. این ویژگی موجب می‌شود که علاوه بر تشخیص وجود ضایعه، محل تقریبی آن نیز مشخص شود و فرآیند بررسی تصویر برای متخصص ساده‌تر گردد[۱۰] .

در برخی کاربردها نیز تعیین مرز دقیق ناحیه آسیب‌دیده اهمیت دارد. در این حالت از روش‌های بخش‌بندی تصویر[۱۳] استفاده می‌شود که به‌جای ارائه یک جعبه محدودکننده، مرز دقیق هر ناحیه را در سطح پیکسل مشخص می‌کنند. اگرچه این روش اطلاعات دقیق‌تری در اختیار قرار می‌دهد، اما معمولاً به داده‌های برچسب‌گذاری‌شده دقیق‌تر، هزینه آماده‌سازی بیشتر و پیچیدگی محاسباتی بالاتری نیاز دارد [۱۱]. بر اساس این تفاوت‌ها، سه وظیفه اصلی تحلیل تصویر را می‌توان به‌صورت خلاصه در جدول ۱-۲ مقایسه کرد.

وظیفه	کاربرد	خروجی
طبقه‌بندی تصویر	تشخیص وجود یا عدم وجود بیماری	کلاس تصویر
آشکارسازی اشیا	تشخیص و مکان‌یابی ضایعات	کلاس و موقعیت مکانی شیء
بخش‌بندی تصویر	اندازه‌گیری و استخراج دقیق ضایعات	مرز دقیق ناحیه

جدول ۱-۲. مقایسه وظایف اصلی تحلیل تصاویر

در حوزه آشکارسازی اشیا، معماری‌های توسعه‌یافته را می‌توان به دو گروه کلی تقسیم کرد. گروه نخست، روش‌های دومرحله‌ای[۱۴] هستند که ابتدا نواحی کاندید را در تصویر استخراج کرده و سپس هر ناحیه را به‌طور جداگانه طبقه‌بندی می‌کنند. خانواده R-CNN از

[۱۱] **Image Classification**
[۱۲] **Bounding Box**
[۱۳] **Image Segmentation**
[۱۴] **Two-stage Detectors**

شناخته‌شده‌ترین نمونه‌های این گروه محسوب می‌شود و معمولاً دقت بالایی دارد، اما به دلیل انجام پردازش در دو مرحله، سرعت اجرای آن نسبتاً پایین‌تر است[۱۲] .

در مقابل، روش‌های تک‌مرحله‌ای[۱۵] فرآیند تشخیص و مکان‌یابی را به‌صورت هم‌زمان انجام می‌دهند و در نتیجه سرعت بیشتری دارند. خانواده‌هایی مانندSSD ، RetinaNetو یولو در این دسته قرار می‌گیرند و به دلیل ایجاد تعادل مناسب میان سرعت و دقت، در بسیاری از کاربردهای عملی مورد استفاده قرار گرفته‌اند [۱۳] [۱٤].

ویژگی	روش‌های تک‌مرحله‌ای	روش‌های دومرحله‌ای
فرآیند تشخیص	یک مرحله	دو مرحله
سرعت اجرا	بیشتر	کمتر
پیچیدگی محاسباتی	کمتر	بیشتر
کاربرد رایج	کاربردهای بلادرنگ و عملی	دقت بالا

جدول ۲-۲. مقایسه کلی روش‌های دومرحله‌ای و تک‌مرحله‌ای آشکارسازی اشیا

با توجه به ماهیت مسئله مورد بررسی در این پژوهش، هدف تنها تعیین وجود شکستگی نیست، بلکه مشخص کردن محل تقریبی آن در تصویر رادیوگرافی نیز اهمیت دارد. ارائه موقعیت شکستگی می‌تواند به پزشک در بررسی سریع‌تر تصویر، افزایش اطمینان تشخیص و کاهش احتمال نادیده‌ماندن ضایعات کمک کند. از سوی دیگر، استفاده از روش‌های بخش‌بندی، اگرچه اطلاعات دقیق‌تری ارائه می‌دهد، مستلزم برچسب‌گذاری بسیار دقیق در سطح پیکسل است؛ در حالی که مجموعه‌داده مورد استفاده در این پژوهش بر پایه جعبه‌های محدودکننده برچسب‌گذاری شده است. بنابراین، آشکارسازی اشیا از نظر نوع داده، هدف پژوهش و نوع خروجی مورد انتظار، مناسب‌ترین چارچوب برای این مطالعه محسوب می‌شود.

در میان روش‌های تک‌مرحله‌ای، خانواده یولو به دلیل ایجاد تعادل مناسب میان دقت، سرعت پردازش و سادگی استقرار، در سال‌های اخیر به یکی از پرکاربردترین چارچوب‌های آشکارسازی اشیا در کاربردهای پزشکی تبدیل شده است. نسخه‌های مختلف این خانواده با هدف بهبود عملکرد، کاهش هزینه محاسباتی و افزایش دقت توسعه یافته‌اند و در بسیاری از مسائل مرتبط با تحلیل تصاویر پزشکی مورد استفاده قرار گرفته‌اند. ازاین‌رو، در بخش بعدی، روند تکامل خانوادهYOLO ، ویژگی‌های معماری آن و دلایل انتخاب این خانواده برای پژوهش حاضر بررسی خواهد شد.

۲-۶ تکامل خانواده مدل‌های یولو و معرفی مدل یولو ۲۶

خانواده یولو یکی از مهم‌ترین چارچوب‌های آشکارسازی اشیا در حوزه بینایی ماشین است که با هدف انجام هم‌زمان طبقه‌بندی و مکان‌یابی اشیا در قالب یک شبکه واحد توسعه یافته است. در این رویکرد، کل تصویر تنها در یک مرحله پردازش می‌شود و مدل

[۱۵] **One-stage Detectors**

به‌صورت هم‌زمان مختصات جعبه‌های محدودکننده، کلاس اشیا و میزان اطمینان هر پیش‌بینی را تولید می‌کند. این رویکرد نسبت به روش‌های دومرحله‌ای، پیچیدگی محاسباتی را کاهش داده و امکان استفاده از آشکارسازی اشیا در کاربردهای بلادرنگ را فراهم ساخته است[۱۵] .

بررسی روند توسعه خانواده یولو نشان می‌دهد که تکامل این مدل‌ها صرفاً به معرفی نسخه‌های جدید محدود نبوده، بلکه بر بهبود مستمر معماری شبکه، راهبردهای آموزش و افزایش کارایی استنتاج متمرکز بوده است. در نسل‌های جدید، هدف اصلی افزایش دقت بدون کاهش سرعت، بهبود استخراج ویژگی در مقیاس‌های مختلف و ساده‌سازی فرآیند آموزش بوده است؛ به‌گونه‌ای که قابلیت استفاده از این خانواده در کاربردهای عملی و صنعتی حفظ شود[۱۵] .

در مقاله مروری YOLOv۸، نویسندگان نشان می‌دهند که حرکت به سمت معماری Anchor-Free، بازطراحی ساختار استخراج ویژگی و استفاده از معماری‌های سبک‌تر، علاوه بر ساده‌تر شدن فرآیند آموزش، موجب بهبود عملکرد مدل در تشخیص اشیا با اندازه‌های مختلف شده است .[۱۵]

در ادامه این روند، Wangو همکاران در مقاله معرفی یولو ۹ مفهوم Programmable Gradient Information (PGI)را با هدف کاهش اتلاف اطلاعات در فرآیند آموزش معرفی کردند. آن‌ها همچنین معماری Generalized Efficient Layer Aggregation Network (GELAN)را برای افزایش بهره‌وری پارامترها و بهبود انتقال ویژگی‌ها ارائه کردند و نتایج آزمایش‌های آن‌ها نشان داد که ترکیب PGI و GELAN بدون افزایش چشمگیر هزینه محاسباتی، عملکرد آشکارسازی را بهبود می‌بخشد .[۱٦]

بر اساس مقاله معرفی یولو ۲۶ ، روند تکامل این خانواده در نسل جدید بیش از آنکه بر تغییرات اساسی در Backbone متمرکز باشد، بر بازطراحی فرآیند آموزش و استنتاج تمرکز دارد. نویسندگان برای این منظور یک معماری دوشاخه (Dual-Head) ارائه کرده‌اند که امکان انجام استنتاج End-to-Endرا بدون نیاز به Non-Maximum Suppression (NMS)فراهم می‌کند. همچنین در این مدل، Distribution Focal Loss (DFL)از بخش پیش‌بینی حذف شده و راهبردهای آموزشی MuSGD، Progressive Loss و Small-Target-Aware Label Assignment (STAL)به‌منظور افزایش سرعت همگرایی، کاهش هزینه آموزش و بهبود تشخیص اشیای کوچک معرفی شده‌اند .[۱۷]

با وجود تفاوت‌های میان نسخه‌های مختلف، ساختار کلی خانواده یولو همچنان بر سه مؤلفه اصلی Backbone، Neckو Headاستوار است. Backbone مسئول استخراج ویژگی‌های تصویری، Neckمسئول ادغام اطلاعات در مقیاس‌های مختلف و Head مسئول پیش‌بینی کلاس، موقعیت مکانی و میزان اطمینان اشیا است. آنچه نسل‌های جدید را از نسخه‌های اولیه متمایز می‌کند، نه تغییر این ساختار کلی، بلکه بهبود نحوه تعامل این مؤلفه‌ها و افزایش بهره‌وری فرآیند آموزش و استنتاج است[۱۷] .

در مقاله یولو ۲۶ ، علاوه بر بهبود عملکرد آشکارسازی اشیا، این چارچوب به‌عنوان یک بستر یکپارچه برای وظایف مختلف بینایی ماشین شامل آشکارسازی اشیا، بخش‌بندی نمونه، برآورد وضعیت بدن، طبقه‌بندی تصویر و آشکارسازی جعبه‌های محدودکننده چرخیده معرفی

شده است. نویسندگان همچنین نسخه YOLOE-۲٦را برای آشکارسازی با واژگان باز (Open-Vocabulary Detection) ارائه کرده‌اند که امکان استفاده از راهنماهای متنی و تصویری را در فرآیند آشکارسازی فراهم می‌کند [۱۷] .

با وجود این پیشرفت‌ها، نویسندگان مقالات مختلف خانواده یولو نیز تأکید کرده‌اند که عملکرد مدل همچنان به کیفیت داده‌های آموزشی، نحوه برچسب‌گذاری، تنوع نمونه‌ها و شرایط استقرار وابسته است. این موضوع در کاربردهای پزشکی اهمیت بیشتری پیدا می‌کند، زیرا وجود ضایعات کوچک، کنتراست پایین تصاویر و تغییرات طبیعی ساختارهای آناتومیکی می‌تواند بر قابلیت تعمیم مدل اثر بگذارد. ازاین‌رو، ارزیابی مستقل هر نسخه بر روی مجموعه‌داده‌های استاندارد پزشکی همچنان ضروری است [۱۶] [۱۷].

بر همین اساس، در پژوهش حاضر یولو ۲۶ به‌عنوان مدل پایه انتخاب شده است. این انتخاب صرفاً بر مبنای جدید بودن نسخه انجام نشده، بلکه بر اساس ویژگی‌هایی نظیر استنتاج End-to-End، حذف وابستگی بهNMS ، بهبود راهبردهای آموزش و توجه ویژه به تشخیص اشیای کوچک صورت گرفته است؛ ویژگی‌هایی که با ماهیت مسئله تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان هم‌خوانی دارند. جزئیات پیاده‌سازی، تنظیمات آموزش و پارامترهای مدل در فصل سوم تشریح خواهند شد.

۲-۷ مجموعه‌داده

یکی از عوامل مؤثر در توسعه و ارزیابی مدل‌های مبتنی بر یادگیری عمیق، دسترسی به مجموعه‌داده‌های استاندارد، باکیفیت و دارای برچسب‌گذاری معتبر است. در حوزه تشخیص شکستگی‌های مچ دست کودکان، محدودیت تعداد مجموعه‌داده‌های عمومی و دشواری تهیه برچسب‌های دقیق توسط متخصصان، همواره یکی از چالش‌های اصلی پژوهشگران بوده است. به همین دلیل، معرفی مجموعه‌داده GRAZPEDWRI-DX گامی مهم در فراهم کردن یک بستر استاندارد برای آموزش و ارزیابی مدل‌های هوش مصنوعی در این حوزه محسوب می‌شود.

این مجموعه توسط پژوهشگران دانشگاه علوم پزشکی گراتس[۱۶] با هدف توسعه و ارزیابی سامانه‌های هوشمند تشخیص شکستگی در کودکان تهیه شده است. این مجموعه‌داده شامل تصاویر رادیوگرافی مچ دست کودکان است که از بیماران واقعی جمع‌آوری شده و پس از طی مراحل کنترل کیفیت، ناشناس‌سازی و بررسی تخصصی، در اختیار جامعه پژوهشی قرار گرفته است. هدف اصلی از انتشار این مجموعه‌داده، فراهم کردن یک مرجع استاندارد برای مقایسه عملکرد روش‌های مختلف یادگیری ماشین و یادگیری عمیق در تشخیص شکستگی‌های مچ دست کودکان است.

بر اساس مقاله معرفی مجموعه‌داده، تصاویر موجود از معاینات واقعی بیماران استخراج شده‌اند و طی فرآیند برچسب‌گذاری، هر تصویر توسط متخصصان رادیولوژی و ارتوپدی بررسی شده است. علاوه بر تعیین وجود یا عدم وجود شکستگی، محل ضایعه نیز با استفاده از

[۱٦] Medical University of Graz

جعبه‌های محدودکننده مشخص شده است. این ویژگی موجب می‌شود که مجموعه‌داده نه‌تنها برای مسائل طبقه‌بندی تصویر، بلکه برای آموزش و ارزیابی مدل‌های آشکارسازی اشیا نیز مناسب باشد.

یکی از مهم‌ترین ویژگی‌های این مجموعه ، کیفیت بالای فرآیند برچسب‌گذاری آن است. در مقاله معرفی این مجموعه‌داده، تأکید شده است که برچسب‌ها بر اساس گزارش‌های بالینی و بازبینی متخصصان تهیه شده‌اند تا احتمال خطاهای ناشی از برچسب‌گذاری به حداقل برسد. وجود چنین فرآیند کنترلی، اعتبار علمی مجموعه‌داده را افزایش داده و امکان استفاده از آن را به‌عنوان یک مرجع پژوهشی فراهم می‌سازد.

علاوه بر کیفیت برچسب‌ها، این مجموعه‌داده از نظر تنوع نمونه‌ها نیز دارای اهمیت است. تصاویر شامل بیماران در گروه‌های سنی مختلف دوران کودکی بوده و انواع گوناگون شکستگی‌های ناحیه مچ دست را پوشش می‌دهد. همچنین تصاویر تحت شرایط واقعی تصویربرداری و با کیفیت‌های متفاوت تهیه شده‌اند؛ موضوعی که باعث می‌شود داده‌ها به شرایط عملی مراکز درمانی نزدیک‌تر باشند و ارزیابی مدل‌ها واقع‌بینانه‌تر انجام شود.

از دیدگاه پژوهشی، این مجموعه چند مزیت مهم دارد. نخست آنکه یک مجموعه‌داده عمومی است و امکان تکرارپذیری آزمایش‌ها و مقایسه مستقیم نتایج پژوهش‌های مختلف را فراهم می‌کند. دوم آنکه تصاویر آن توسط متخصصان پزشکی برچسب‌گذاری شده‌اند و بنابراین از اعتبار بالینی مناسبی برخوردار هستند. سوم آنکه وجود برچسب‌های مکانی در قالب جعبه‌های محدودکننده، این مجموعه‌داده را برای ارزیابی مدل‌های آشکارسازی اشیا، از جمله خانوادهYOLO ، مناسب می‌سازد.

با وجود این مزایا، این مجموعه‌داده نیز همانند سایر مجموعه‌داده‌های پزشکی دارای محدودیت‌هایی است. این مجموعه‌داده تنها بر تصاویر رادیوگرافی مچ دست کودکان متمرکز است و بنابراین نتایج حاصل از مدل‌های آموزش‌دیده بر روی آن را نمی‌توان بدون ارزیابی مجدد به سایر اندام‌ها، گروه‌های سنی یا روش‌های تصویربرداری پزشکی تعمیم داد. همچنین، عملکرد مدل‌های آموزش‌دیده ممکن است در مراکز درمانی با تجهیزات تصویربرداری متفاوت یا جمعیت‌های آماری دیگر نیازمند اعتبارسنجی مجدد باشد. این محدودیت‌ها در مقاله معرفی مجموعه‌داده نیز مورد توجه قرار گرفته‌اند و ضرورت ارزیابی مستقل مدل‌ها در شرایط مختلف را نشان می‌دهند.

با توجه به ویژگی‌های یادشده، مجموعه‌داده دانشگاه گراتس بستر مناسبی برای ارزیابی عملکرد مدل‌های آشکارسازی اشیا در تشخیص شکستگی‌های مچ دست کودکان فراهم می‌کند. استفاده از تصاویر واقعی بیماران، برچسب‌گذاری تخصصی، دسترس‌پذیری عمومی و امکان مقایسه نتایج با سایر پژوهش‌ها، از مهم‌ترین دلایل انتخاب این مجموعه‌داده در پژوهش حاضر است. بر این اساس، تمامی مراحل آموزش، اعتبارسنجی و ارزیابی مدل یولو ۲۶ در این پژوهش بر پایه داده‌های این مجموعه‌داده انجام شده است. جزئیات نحوه تقسیم داده‌ها، پیش‌پردازش تصاویر و تنظیمات آموزشی در فصل سوم تشریح خواهد شد.

۲-۸ معیارهای ارزیابی عملکرد مدل

ارزیابی عملکرد مدل‌های یادگیری عمیق، یکی از مراحل اساسی در سنجش قابلیت استفاده آن‌ها در کاربردهای واقعی است. در مسئله آشکارسازی اشیا، هدف تنها تشخیص وجود یک شیء نیست، بلکه مدل باید محل دقیق آن را نیز با دقت مناسبی مشخص کند. به همین دلیل، معیارهای مورد استفاده برای ارزیابی مدل‌های آشکارسازی اشیا با معیارهای رایج در مسائل طبقه‌بندی تفاوت دارند و علاوه بر صحت تشخیص، کیفیت مکان‌یابی اشیا را نیز در نظر می‌گیرند. در پژوهش حاضر، ارزیابی عملکرد مدل یولو ۲۶ بر اساس معیارهای استاندارد مورد استفاده در ادبیات آشکارسازی اشیا انجام شده است.

۲-۸-۱ مفاهیم پایه ارزیابی

در ارزیابی مدل‌های آشکارسازی اشیا، هر پیش‌بینی مدل با برچسب مرجع[۱۷] مقایسه می‌شود. بر این اساس، نتایج پیش‌بینی را می‌توان در سه گروه اصلی طبقه‌بندی کرد:

- مثبت صحیح[۱۸] یا TP : مدل شکستگی را به‌درستی تشخیص داده و محل پیش‌بینی‌شده با محل واقعی هم‌پوشانی قابل قبول دارد.
- مثبت کاذب[۱۹] : مدل وجود شکستگی را گزارش کرده است، اما پیش‌بینی نادرست بوده یا محل آن با برچسب واقعی تطابق کافی ندارد.
- منفی کاذب[۲۰]: شکستگی واقعی در تصویر وجود داشته است، اما مدل موفق به تشخیص آن نشده است.

در مسائل آشکارسازی اشیا، برخلاف بسیاری از مسائل طبقه‌بندی، معیار [۲۱] (TN) معمولاً در ارزیابی عملکرد نقشی ندارد؛ زیرا تعداد نواحی فاقد شیء در تصویر بسیار زیاد بوده و محاسبه آن اطلاعات مفیدی درباره عملکرد مدل ارائه نمی‌کند. به همین دلیل، معیارهای ارزیابی آشکارسازی عمدتاً بر پایه TP ، FP و FN تعریف می‌شوند.

۲-۸-۲ معیار هم‌پوشانی[۲۲]

پیش از محاسبه معیارهای ارزیابی، لازم است مشخص شود که آیا محل پیش‌بینی‌شده توسط مدل با محل واقعی شیء تطابق کافی دارد یا خیر. برای این منظور از معیار IoU استفاده می‌شود. این معیار نسبت مساحت اشتراک جعبه محدودکننده پیش‌بینی‌شده و جعبه مرجع به مساحت اجتماع آن‌ها را نشان می‌دهد و به صورت زیر تعریف می‌شود:

---

[۱۷] Ground Truth
[۱۸] **True Positive**
[۱۹] **False Positive**
[۲۰] **False Negative**
[۲۱] **True Negative**
[۲۲] **Intersection over Union**

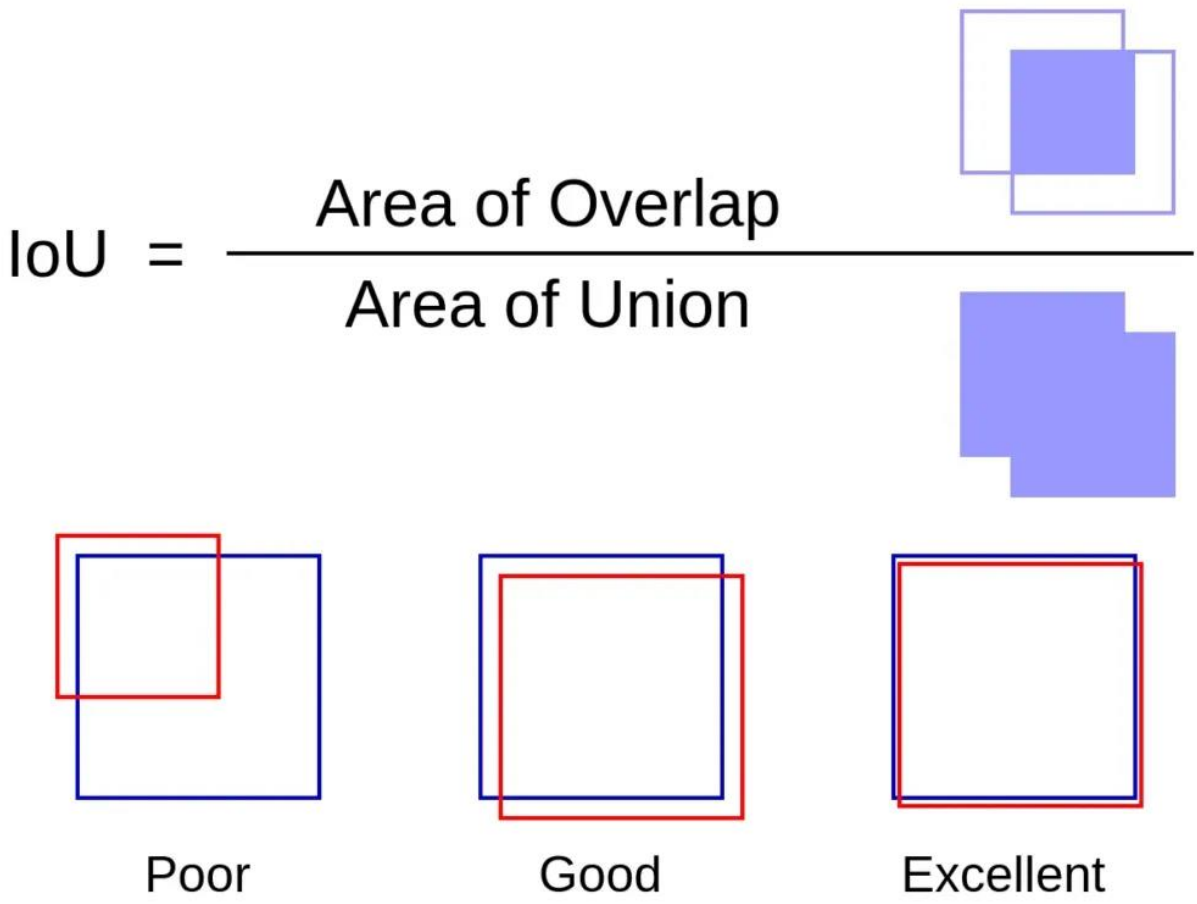
$$IoU = \frac{| B_{pred} \cap B_{gt} |}{| B_{pred} \cup B_{gt} |}$$

که در آن:

- B_{pred}جعبه محدودکننده پیش‌بینی‌شده توسط مدل،
- B_{gt}جعبه محدودکننده مرجع

را نشان می‌دهند.

مقدار IoU بین صفر و یک قرار می‌گیرد. هرچه این مقدار به یک نزدیک‌تر باشد، محل پیش‌بینی‌شده تطابق بیشتری با محل واقعی شیء دارد. در بسیاری از ارزیابی‌های استاندارد، از جمله ارزیابی مدل‌های خانوادهYOLO ، آستانه IoU=۰٫۵۰به‌عنوان حداقل شرط پذیرش یک تشخیص صحیح در نظر گرفته می‌شود.



شکل ۱-۲: نمایش مفهوم Intersection over Union پیشنهاد می‌شود.

۲-۸-۳ معیار دقت مثبت

دقت مثبت (Precision) نشان می‌دهد از میان تمام شکستگی‌هایی که مدل گزارش کرده است، چه نسبتی واقعاً صحیح بوده‌اند.

رابطه این معیار به صورت زیر است:

$$Precision = \frac{TP}{TP + FP}$$

مقدار بالای دقت مثبت بیانگر آن است که مدل تعداد کمی هشدار اشتباه تولید کرده و بیشتر پیش‌بینی‌های آن صحیح هستند. در کاربردهای پزشکی، افزایش دقت مثبت موجب کاهش تشخیص‌های مثبت کاذب و جلوگیری از انجام بررسی‌های غیرضروری می‌شود.

۲-۸-۴ معیار فراخوانی

فراخوانی (Recall) توانایی مدل در یافتن تمامی شکستگی‌های موجود را اندازه‌گیری می‌کند و به صورت زیر تعریف می‌شود:

$$Recall = \frac{TP}{TP + FN}$$

هرچه مقدار فراخوانی بیشتر باشد، احتمال نادیده‌ماندن شکستگی‌های واقعی کمتر خواهد بود. در سامانه‌های کمک‌تشخیصی پزشکی، فراخوانی اهمیت ویژه‌ای دارد؛ زیرا عدم تشخیص یک شکستگی واقعی می‌تواند پیامدهای بالینی نامطلوبی به همراه داشته باشد.

۲-۸-۵ معیار F۱-Score

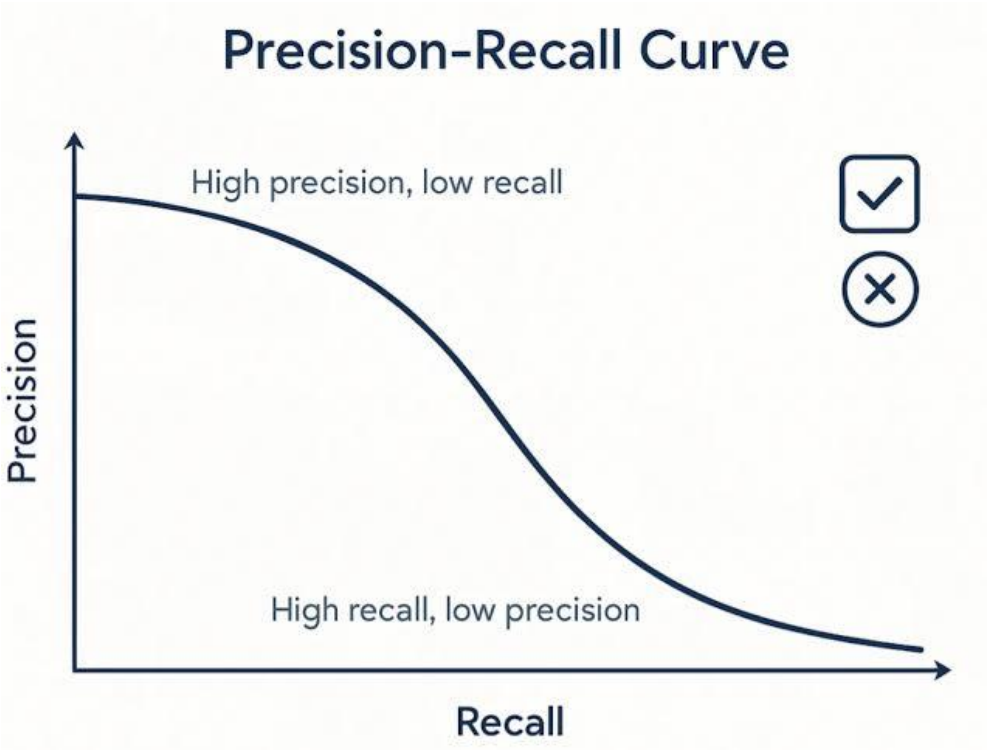
در بسیاری از موارد، افزایش دقت مثبت موجب کاهش فراخوانی و بالعکس می‌شود. برای ایجاد تعادل میان این دو معیار، از امتیاز F۱ استفاده می‌شود که میانگین هارمونیک دقت مثبت و فراخوانی است.

$$F1 = \frac{۲ \times Precision \times Recall}{Precision + Recall}$$

مقدار بالای امتیاز F۱ نشان‌دهنده عملکرد متعادل مدل در کاهش خطاهای مثبت کاذب و منفی کاذب است.

۶-۸-۲ میانگین دقت

در ارزیابی مدل‌های آشکارسازی اشیا، تنها محاسبه دقت مثبت و فراخوانی در یک آستانه مشخص کافی نیست. به همین دلیل، با تغییر آستانه اطمینان مدل، منحنی دقت مثبت–فراخوانی رسم شده و میانگین دقت مثبت (AP) به‌عنوان مساحت زیر این منحنی محاسبه می‌شود. AP عملکرد مدل را در تمامی سطوح اطمینان پیش‌بینی در نظر می‌گیرد و یکی از مهم‌ترین معیارهای ارزیابی مدل‌های آشکارسازی اشیا محسوب می‌شود.



شکل ۲-۲: نمونه منحنی دقت مثبت–فراخوانی

۷-۸-۲ میانگین دقت مثبت(mAP)

از آنجا که عملکرد مدل ممکن است در کلاس‌های مختلف یا آستانه‌های متفاوت IoU تغییر کند، در ارزیابی استاندارد از معیار دقت مثبت[۲۳] (mAP) استفاده می‌شود.

در پژوهش حاضر دو معیار رایج زیر مورد استفاده قرار گرفته‌اند:

- mAP@۰٫۵۰ که در آن یک پیش‌بینی زمانی صحیح محسوب می‌شود که مقدار IoU حداقل برابر ۰.۵۰ باشد.

[۲۳] **Mean Average**

- mAP@۰,۵۰:۰,۹۵ که میانگین مقدار AP را در آستانه‌های مختلف IoU از ۰.۵۰ تا ۰.۹۵ با گام ۰.۰۵ محاسبه می‌کند. این معیار ارزیابی سخت‌گیرانه‌تری ارائه می‌دهد و در بسیاری از مجموعه‌داده‌های استاندارد، از جمله مجموعه‌داده‌های مورد استفاده در رقابت‌های بینایی ماشین، به‌عنوان معیار اصلی گزارش عملکرد مدل‌ها به کار می‌رود.

به دلیل در نظر گرفتن هم‌زمان کیفیت تشخیص و دقت مکان‌یابی، mAP به‌عنوان مهم‌ترین شاخص مقایسه عملکرد مدل‌های آشکارسازی اشیا شناخته می‌شود.

۲-۸-۸ جمع‌بندی

معیارهای معرفی‌شده در این بخش، چارچوب استاندارد ارزیابی مدل‌های آشکارسازی اشیا را تشکیل می‌دهند و در اغلب پژوهش‌های مرتبط با خانواده یولو مورد استفاده قرار می‌گیرند. در پژوهش حاضر نیز عملکرد مدل یولو ۲۶ بر اساس معیارهای دقت مثبت، فراخوانی، F۱-Score، mAP@۰,۵۰ و mAP@۰,۵۰:۰,۹۵ ارزیابی خواهد شد. نتایج حاصل از این ارزیابی‌ها در فصل پنجم ارائه و تحلیل می‌شوند و مبنای مقایسه عملکرد مدل با مطالعات پیشین قرار خواهند گرفت.

۲-۹ پژوهش‌های پیشین

پژوهش درباره تشخیص خودکار شکستگی مچ دست کودکان در سال‌های اخیر، به‌ویژه پس از انتشار مجموعه‌داده عمومی دانشگاه گراتس، گسترش یافته است. مطالعات این حوزه را می‌توان در سه مسیر اصلی دسته‌بندی کرد: نخست، ایجاد بستر داده‌ای استاندارد؛ دوم، ارزیابی نسل‌های مختلف مدل‌های آشکارسازی؛ و سوم، اصلاح معماری، راهبرد آموزش یا طراحی مجموعه آزمون برای نزدیک‌تر کردن ارزیابی‌ها به شرایط واقعی. این دسته‌بندی نشان می‌دهد که پیشرفت حوزه صرفاً حاصل جایگزینی یک نسخه یولو با نسخه جدیدتر نبوده، بلکه به کیفیت داده، نحوه تقسیم مجموعه‌داده، اندازه تصویر، افزایش داده، سازوکار تخصیص برچسب و طراحی اجزای شبکه نیز وابسته بوده است.

۲-۹-۱ شکل‌گیری بستر استاندارد پژوهش

نقطه عطف اصلی این حوزه، انتشار مجموعه‌داده مذکور توسط Nagy و همکاران در سال ۲۰۲۲ بود. این مجموعه شامل ۱۰،۶۴۳ مطالعه رادیوگرافی، ۲۰،۳۲۷ تصویر متعلق به ۶،۰۹۱ بیمار کودک و نوجوان، ۷۴،۴۵۹ برچسب تصویری و ۶۷،۷۷۱ شیء نشانه‌گذاری‌شده است. تصاویر عمدتاً نماهای پس‌خلفی و جانبی را پوشش می‌دهند و ضایعات با خطوط، چندضلعی‌ها یا جعبه‌های محدودکننده توسط متخصصان رادیولوژی کودکان نشانه‌گذاری شده‌اند. انتشار عمومی این مجموعه‌داده، امکان آموزش و مقایسه مدل‌ها را بر یک بستر مشترک فراهم کرد؛ هرچند داده‌ها از یک مرکز درمانی جمع‌آوری شده‌اند و تعمیم نتایج به جمعیت‌ها، تجهیزات و مراکز دیگر نیازمند اعتبارسنجی خارجی است .[۱۸]

اهمیت این مجموعه‌داده فقط به تعداد تصاویر محدود نیست. وجود برچسب‌های مکانی، آن را برای مسئله آشکارسازی اشیا مناسب ساخته و امکان بررسی هم‌زمان تشخیص و مکان‌یابی ضایعات را فراهم کرده است. بااین‌حال، مقاله معرفی مجموعه‌داده یک مدل مرجع نهایی برای تشخیص شکستگی ارائه نمی‌کند؛ بنابراین، پژوهش‌های بعدی وظیفه تعیین خط پایه، اصلاح فرایند آموزش و مقایسه آشکارسازهای مختلف را بر عهده گرفته‌اند.

۲-۹-۲ ارزیابی و بهینه‌سازی YOLOv۷

یکی از نخستین مطالعات نظام‌مند مبتنی بر مجموعه‌داده GRAZPEDWRI-DX توسط مایر و همکاران انجام شد. پژوهشگران ابتدا یک مدل پایه مبتنی بر یولو ۷ را آموزش دادند و سپس تأثیر روش‌های آماده‌سازی داده، اندازه تصویر ورودی و تنظیمات معماری و آموزش را بر عملکرد مدل بررسی کردند. نتایج نشان داد که بهینه‌سازی فرآیند آماده‌سازی داده‌ها می‌تواند میانگین دقت در آستانه هم‌پوشانی ۵۰ درصد را تا ۲۵٫۵۱ درصد و میانگین دقت در بازه آستانه‌های هم‌پوشانی ۵۰ تا ۹۵ درصد را تا ۳۹٫۷۸ درصد بهبود دهد. همچنین، اعمال تغییرات در معماری مدل و تنظیمات آموزش نیز موجب افزایش بیشتر عملکرد شد. در ادامه، پژوهشگران با استفاده از روش Grad-CAM الگوی توجه مدل را تحلیل کردند و نشان دادند مدل‌هایی که عملکرد بهتری دارند، در فرآیند تصمیم‌گیری از نواحی گسترده‌تری از تصویر بهره می‌گیرند که این موضوع بیانگر استخراج مؤثرتر ویژگی‌های مرتبط با شکستگی است [۱۹].

اهمیت این پژوهش در آن است که نشان داد عملکرد مدل تنها تابع انتخاب نام معماری نیست و تصمیم‌هایی مانند وضوح ورودی، پاک‌سازی برچسب‌ها و پیکربندی آموزش می‌توانند نتیجه را به‌طور اساسی تغییر دهند. از سوی دیگر، تعدد تغییرات آزمایشی و وابستگی نتایج به یک مجموعه‌داده واحد، مقایسه مستقیم آن با مطالعاتی که تقسیم داده یا پروتکل آموزش متفاوتی دارند را دشوار می‌سازد. این محدودیت، ضرورت گزارش دقیق تقسیم داده و تنظیمات آزمایش را برجسته می‌کند.

۲-۹-۳ استفاده از YOLOv۸ و راهبردهای افزایش داده

Ju و Cai در سال ۲۰۲۳ عملکرد گونه‌های مختلف YOLOv۸ را بر روی GRAZPEDWRI-DX بررسی کردند و برای افزایش تنوع نمونه‌های آموزشی از راهبردهای افزایش داده استفاده کردند. مدل بهینه‌شده آنان به $mAP@0.5 = ۰.۶۳۸$ دست یافت؛ درحالی‌که برای نسخه بهبودیافته YOLOv۷ و YOLOv۸ پایه به‌ترتیب مقادیر ۰٫۶۳۴ و ۰٫۶۳۶ گزارش شد. آنان علاوه بر مدل، یک رابط کاربردی برای نمایش محل شکستگی نیز ارائه کردند .[۲۰]

اگرچه برتری عددی مدل بهینه‌شده نسبت به خط پایه محدود بود، این مطالعه نشان داد که افزایش داده و انتخاب اندازه مناسب مدل می‌تواند بدون طراحی یک معماری کاملاً جدید، عملکرد را بهبود دهد. بااین‌حال، تقسیم آموزش، اعتبارسنجی و آزمون در این پژوهش به‌صورت تصادفی انجام شده و نویسندگان تصریح کرده‌اند که تقسیم ایجادشده بازتولیدپذیر نیست. این موضوع مقایسه نتایج آن با پژوهش‌هایی که از تقسیم‌های متفاوت استفاده کرده‌اند را با احتیاط همراه می‌کند.

در ادامه همین مسیر، مطالعاتی با افزودن ماژول‌های توجه به YOLOv۸ انجام شد. برای نمونه، مدل YOLOv۸-GCبا افزودن بلوک زمینه سراسری، مقدار $mAP@0.5$را از ۶۳٫۵۸ به ۶۶٫۳۲ درصد افزایش داد. مدل YOLOv۸-ResCBAMنیز با ترکیب ماژول توجه کانالی و مکانی، $mAP@0.5$را از حدود ۶۳٫۶ به ۶۵٫۸ درصد رساند. این نتایج نشان می‌دهند که تقویت مدل‌سازی زمینه و هدایت توجه شبکه می‌تواند در آشکارسازی خطوط شکستگی ظریف مؤثر باشد؛ اما افزایش پیچیدگی معماری و ارزیابی عمدتاً درون‌داده‌ای، همچنان نیاز به سنجش تعمیم‌پذیری در داده‌های مستقل را باقی می‌گذارد [۲۱] .[۲۲]

۲-۹-۴ مقایسه آشکارسازهای تک‌مرحله‌ای و دومرحله‌ای

در یک مطالعه مقایسه‌ای، احمد و همکاران نسخه‌های مختلفYOLOv۵ ، YOLOv٦، YOLOv۷و YOLOv۸ را در کنار Faster R-CNNبررسی کردند. بر اساس نتایج گزارش‌شده، مدل‌های تک‌مرحله‌ای یولو در این مسئله عملکرد رقابتی‌تری از Faster R-CNN نشان دادند. YOLOv۸m برای کلاس شکستگی حساسیت ۰٫۹۲ و mAP برابر ۰٫۹۵ و YOLOv۸x در ارزیابی همه کلاس‌ها mAP برابر ۰٫۷۷ به دست آورد. این مطالعه از مناسب بودن آشکارسازهای تک‌مرحله‌ای برای ایجاد تعادل میان دقت و کارایی حمایت می‌کند، ولی تفاوت میان معیارهای گزارش‌شده برای کلاس شکستگی و کل کلاس‌ها، مقایسه اعداد را بدون توجه به دامنه ارزیابی نامعتبر می‌سازد .[۲۳]

این نکته برای پژوهش حاضر اهمیت دارد: نتایجی که تنها برای کلاس «شکستگی» محاسبه شده‌اند، نباید مستقیماً با mAP کل کلاس‌های مجموعه‌داده مقایسه شوند. همچنین اختلاف در گونه مدل، وضوح ورودی، تقسیم بیماران و آستانه‌های ارزیابی می‌تواند اختلاف عملکرد ایجاد کند؛ بنابراین، مقایسه منصفانه مستلزم یک پروتکل مشترک است.

۲-۹-۵ به‌کارگیری یولو ۹ و یولو ۱۰

Chienو همکاران در سال ۲۰۲۴ یولو ۹ را برای تشخیص شکستگی در همین دیتاست به کار گرفتند. یولو ۹ با تکیه بر Programmable Gradient Information (PGI)و معماری GELANبرای کاهش اتلاف اطلاعات در شبکه‌های عمیق طراحی شده است. در مطالعه کاربردی شکستگی، مقدار $mAP@0.5:0.95$از ۴۲٫۱۶ به ۴۳٫۷۳ درصد افزایش یافت؛ یعنی بهبود نسبی ۳٫۷ درصدی نسبت به خط مبنای گزارش‌شده .[۲٤]

اهمیت یولو ۹ در این حوزه فقط به افزایش یک معیار محدود نیست؛ این مدل جهت پژوهش را از افزایش داده صرف، به سمت اصلاح جریان اطلاعات و آموزش شبکه هدایت کرد. بااین‌حال، افزایش عملکرد گزارش‌شده همچنان در همان مجموعه‌داده و بدون اعتبارسنجی خارجی به دست آمده است. افزون بر آن، مقاله اصلی یولو ۹ سازوکار PGI و GELAN را روی MS COCO اعتبارسنجی کرده است و کارایی آن در تصویربرداری پزشکی باید مستقلاً سنجیده شود.

پس از آن، Ahmedو Manaf گونه‌های مختلف یولو ۱۰ را با تمرکز بر تخصیص دوگانه برچسب و استنتاج بدون NMS بررسی کردند. بهترین مدل آنان $mAP@0.5:0.95 = 51.9\%$ را گزارش کرد که در مقایسه با خط مبنای یولو ۹ مورد استفاده در همان مطالعه، ۸٫۶ واحد درصد افزایش داشت [۲۵]. این نتیجه ظرفیت مدل‌های انتهابه‌انتها را نشان می‌دهد، اما مقاله به‌صورت پیش‌انتشار منتشر شده و نتایج آن باید با توجه به جزئیات تقسیم داده و پروتکل آزمایش تفسیر شود.

۶-۹-۲ نسل‌های جدیدتر و تمرکز بر تعمیم‌پذیری

از سال ۲۰۲۵ به بعد، پژوهش‌ها به سمت استفاده از YOLO۱۱ و نسخه‌های بهبودیافته آن سوق پیدا کرده‌اند. برای نمونه، در یکی از این مطالعات، یک مدل بهبودیافته مبتنی بر YOLO۱۱ و تابع زیان Focaler-MPDIoU توسعه یافت که نسبت به نسخه پایه YOLO۱۱، ۳٫۲ درصد افزایش دقت، ۱٫۶ درصد افزایش فراخوانی، ۱٫۸ درصد افزایش میانگین دقت در آستانه هم‌پوشانی ۵۰ درصد و ۳٫۲ درصد افزایش میانگین دقت در بازه آستانه‌های هم‌پوشانی ۵۰ تا ۹۵ درصد را گزارش کرده است [۲۶].

در مطالعه‌ای دیگر، مدل FracDet-v۱۱ با بهره‌گیری از سازوکارهای استخراج ویژگی چندمقیاسی، مکانیزم توجه و تبدیل موجک، بر روی مجموعه‌داده GRAZPEDWRI-DX به دقت ۷۳٫۹ درصد و میانگین دقت در آستانه هم‌پوشانی ۵۰ درصد برابر با ۶۴٫۸ درصد دست یافت. یکی از مهم‌ترین ویژگی‌های این پژوهش، ارزیابی مدل بر روی مجموعه‌داده خارجی FracAtlas بود که در آن میانگین دقت در آستانه هم‌پوشانی ۵۰ درصد برابر با ۴۷٫۹ درصد گزارش شد. کاهش عملکرد میان داده‌های داخلی و خارجی نشان می‌دهد که دستیابی به نتایج مطلوب بر روی یک مجموعه‌داده، لزوماً به معنای قابلیت تعمیم مناسب مدل در کاربردهای بالینی نیست [۲۷].

در پژوهشی دیگر، ترکیب شبکه ResNet۵۰ با YOLO۱۱ مورد بررسی قرار گرفت و مقادیر بالایی برای میانگین دقت در آستانه هم‌پوشانی ۵۰ درصد گزارش شد. با این حال، اختلاف قابل‌توجه این نتایج با بسیاری از مطالعات پیشین نشان می‌دهد که پیش از هرگونه مقایسه، عواملی نظیر تعداد و نوع کلاس‌های مورد ارزیابی، روش تقسیم داده‌ها، شیوه پاک‌سازی داده‌ها و احتمال وجود هم‌پوشانی بیماران میان مجموعه‌های آموزش و آزمون باید به‌دقت بررسی شوند [۲۸].

اهمیت طراحی صحیح مجموعه آزمون نیز در مطالعه‌ای مستقل مورد تأکید قرار گرفت. در این پژوهش، مدل‌های EfficientNet و YOLO۱۱ بر روی دو مجموعه آزمون تصادفی و متوازن ارزیابی شدند. نتایج نشان داد عملکرد مدل‌ها در مجموعه آزمون متوازن که شامل نمونه‌های دشوارتر بود، به‌طور معناداری کاهش یافت. پژوهشگران نتیجه گرفتند که استفاده از مجموعه‌های آزمون تصادفی می‌تواند عملکرد واقعی مدل را بیش‌برآورد کند و طراحی و استانداردسازی مجموعه آزمون، نقش تعیین‌کننده‌ای در ارزیابی قابلیت استفاده بالینی مدل‌های هوش مصنوعی دارد [۲۹].جدول ۴-۲ ـ مقایسه پژوهش‌های شاخص

مرجع	رویکرد	نتیجه شاخص گزارش‌شده	داده و دامنه ارزیابی	نکته یا محدودیت اصلی
Nagyو همکاران[۱۸]	ایجاد مجموعه‌داده	انتشار GRAZPEDWRI-DX با ۶۷٬۷۷۱ شیء نشانه‌گذاری‌شده	۲۰٬۳۲۷ تصویر، ۶٬۰۹۱ بیمار	تک‌مرکزی و بدون مدل مرجع نهایی
Mayerو همکاران[۱۹]	YOLOv۷و بهینه‌سازی داده/مدل	بهبود قابل‌توجه mAP با پیش‌پردازش و تنظیم مدل	GRAZPEDWRI-DX	وابستگی زیاد نتیجه به پروتکل آماده‌سازی و آموزش
Ju و Cai [۲۰]	YOLOv۸و افزایش داده	$mAP@0.5 = ۰.۶۳۸$	GRAZPEDWRI-DX	تقسیم تصادفی و غیرقابل‌بازتولید داده
Ju و همکاران [۲۱]	YOLOv۸-GC	افزایش $mAP@0.5$از ۶۳٫۵۸ به ۶۶٫۳۲٪	GRAZPEDWRI-DX	نبود اعتبارسنجی خارجی
Chienو همکاران[۲٤]	یولو ۹، PGIو GELAN	$mAP@0.5:0.95 = ۴۳.۷۳٪$	GRAZPEDWRI-DX	ارزیابی درون یک مجموعه‌داده
Ahmedو Manaf [۲۵]	یولو ۱۰ و تخصیص دوگانه	$mAP@0.5:0.95 = ۵۱.۹٪$	GRAZPEDWRI-DX	پیش‌انتشار و حساس به پروتکل مقایسه
مدل بهبودیافته YOLO۱۱ [۲٦]	Focaler-MPDIoU	بهبود هم‌زمان دقت مثبت، فراخوانی وmAP	GRAZPEDWRI-DX	ارزیابی عمدتاً درون‌داده‌ای
FracDet-v۱۱ [۲۷]	توجه چندمقیاسی و موجک	$mAP@0.5 =$ ۶۴.۸٪داخلی و ۴۷٫۹٪ خارجی	GRAZPEDWRI-DX وFracAtlas	افت محسوس در داده خارجی
مطالعه طراحی مجموعه آزمون [۲۹]	EfficientNetو YOLO۱۱	افت معنادار عملکرد در آزمون دشوار و متوازن	دو آزمون تصادفی و متوازن	نشان‌دهنده حساسیت نتایج به ترکیب آزمون

۷-۹-۲ جمع‌بندی

بررسی مطالعات نشان می‌دهد که ادبیات این حوزه از یک خط سیر مشخص پیروی کرده است. در مرحله نخست، انتشار داده های دانشگاه گراتس، امکان ارزیابی استاندارد مدل‌ها را فراهم کرد. در مرحله بعد، YOLOv۷و YOLOv۸ نشان دادند که پیش‌پردازش، وضوح تصویر و افزایش داده می‌توانند به‌اندازه انتخاب معماری مؤثر باشند. سپس یولو ۹ و یولو ۱۰ توجه را به جریان گرادیان، تخصیص برچسب و استنتاج انتهابه‌انتها معطوف کردند. مطالعات جدیدتر نیز بر توجه چندمقیاسی، زیان‌های مکان‌یابی، سبک‌سازی و اعتبارسنجی خارجی تمرکز کرده‌اند.

بااین‌حال، مقایسه مستقیم مقادیر گزارش‌شده در مطالعات مختلف با چند مانع جدی مواجه است. تقسیم‌های متفاوت داده، احتمال جداسازی در سطح تصویر به‌جای بیمار، تفاوت در تعداد کلاس‌های ارزیابی‌شده، اندازه ورودی، پاک‌سازی برچسب‌ها و معیار انتخاب

بهترین epoch می‌توانند اعداد نهایی را تغییر دهند. مطالعه مربوط به ترکیب مجموعه آزمون نیز به‌صورت تجربی نشان داده است که نمونه‌گیری تصادفی ممکن است عملکرد را بیش‌برآورد کند. بنابراین، صرف دستیابی به mAP بالاتر برای اثبات برتری یک معماری کافی نیست و پروتکل ارزیابی باید هم‌زمان بررسی شود.

از منظر بالینی نیز بیشتر مطالعات بر ارزیابی درون همین دیتاست متمرکز بوده‌اند و تنها تعداد محدودی اعتبارسنجی خارجی یا مقایسه با شرایط دشوارتر انجام داده‌اند. افت FracDet-v۱۱ روی FracAtlas نشان می‌دهد که تغییر مرکز، جمعیت و ویژگی‌های تصویربرداری می‌تواند عملکرد را به‌طور محسوسی کاهش دهد. در نتیجه، مدل‌های این حوزه در وضعیت فعلی باید به‌عنوان سامانه‌های پژوهشی یا کمک‌تشخیصی بالقوه در نظر گرفته شوند، نه ابزارهای آماده استقرار بالینی.

سرانجام، جست‌وجوی منابع مرتبط نشان داد که نسل‌های یولو تا YOLO۱۱ بر روی دیتاست یاد شده بررسی شده‌اند، اما مطالعه منتشرشده و مستقلی که یولو ۲۶ را برای تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان روی این مجموعه‌داده ارزیابی کرده باشد، یافت نشد. مقاله معرفی یولو ۲۶ ویژگی‌هایی مانند سر دوشاخه بدونDFL ، استنتاج انتهابه‌انتها بدونNMS ، بهینه‌سازMuSGD ، Progressive Lossو STAL را ارائه می‌کند؛ ولی نتایج اصلی آن بر مجموعه‌داده‌های عمومی بینایی ماشین گزارش شده‌اند، نه بر تصاویر رادیوگرافی کودکان. بنابراین، انتقال این مزایا به حوزه پزشکی هنوز نیازمند ارزیابی تجربی مستقل است.

این تحلیل، زمینه لازم برای استخراج دقیق خلأ پژوهشی و تبیین جایگاه پژوهش حاضر را در بخش بعد فراهم می‌کند.

۲-۱۰ جمع‌بندی فصل و تبیین خلأ پژوهشی

در این فصل، مبانی نظری و پیشینه پژوهش مرتبط با تشخیص خودکار شکستگی‌های مچ دست کودکان بررسی شد. ابتدا ویژگی‌های بالینی شکستگی‌های این ناحیه و اهمیت تشخیص دقیق و به‌موقع آن‌ها در تصاویر رادیوگرافی تشریح گردید. سپس جایگاه روش‌های مبتنی بر یادگیری عمیق، به‌ویژه آشکارسازی اشیا، در توسعه سامانه‌های کمک‌تشخیصی مورد بررسی قرار گرفت. همچنین روند تکامل خانواده مدل‌هایYOLO ، ویژگی‌های مجموعه‌داده استاندارد GRAZPEDWRI-DXو معیارهای متداول ارزیابی عملکرد مدل‌های آشکارسازی اشیا معرفی شدند. این مباحث نشان دادند که استفاده از مدل‌های آشکارسازی اشیا می‌تواند علاوه بر تشخیص وجود شکستگی، محل تقریبی آن را نیز مشخص کند و از این نظر نسبت به روش‌های صرفاً طبقه‌بندی، تناسب بیشتری با نیازهای سامانه‌های کمک‌تشخیصی دارد.

بررسی مطالعات پیشین نشان داد که پس از انتشار مجموعه‌داده GRAZPEDWRI-DX، پژوهش‌های متعددی با استفاده از نسخه‌های مختلف خانواده یولو برای تشخیص شکستگی‌های مچ دست کودکان انجام شده است. در این مطالعات، علاوه بر اصلاح معماری مدل، راهکارهایی نظیر افزایش داده، بهینه‌سازی فرآیند آموزش، استفاده از ماژول‌های توجه، طراحی توابع زیان جدید و بهبود روش‌های تخصیص برچسب به کار گرفته شده‌اند. نتایج این پژوهش‌ها نشان می‌دهد که چنین اصلاحاتی می‌توانند عملکرد مدل را بهبود دهند، اما

میزان این بهبود به عوامل متعددی از جمله کیفیت داده‌های آموزشی، نحوه تقسیم مجموعه‌داده، اندازه تصاویر ورودی، تنظیمات آموزش و معیارهای ارزیابی وابسته است. ازاین‌رو، مقایسه مستقیم نتایج مطالعات مختلف بدون توجه به تفاوت پروتکل‌های آزمایش، همواره امکان‌پذیر نیست.

از سوی دیگر، مرور ادبیات نشان داد که بخش عمده پژوهش‌های منتشرشده بر ارزیابی نسخه‌های YOLOv۷، YOLOv۸، یولو ۹، یولو ۱۰ و در مطالعات جدیدتر YOLO۱۱متمرکز بوده‌اند. هر یک از این مطالعات تلاش کرده‌اند با اصلاح بخشی از معماری یا فرآیند آموزش، عملکرد مدل را بهبود دهند. با وجود این پیشرفت‌ها، اعتبارسنجی خارجی مدل‌ها همچنان محدود است و اغلب نتایج بر روی همان مجموعه‌داده‌ای گزارش شده‌اند که مدل بر اساس آن آموزش دیده است. این موضوع نشان می‌دهد که تعمیم‌پذیری مدل‌های موجود به داده‌های جدید یا شرایط متفاوت تصویربرداری همچنان یکی از چالش‌های مهم این حوزه محسوب می‌شود.

مطالعه منابع علمی انجام‌شده در این پژوهش همچنین نشان داد که یولو ۲۶ به‌عنوان نسل جدید خانوادهYOLO ، گام جدیدی در توسعه مدل‌های آشکارسازی اشیا برداشته است. با این حال، نتایج ارائه‌شده در مقاله معرفی این مدل عمدتاً بر روی مجموعه‌داده‌های عمومی بینایی ماشین گزارش شده‌اند و شواهد مستقلی درباره عملکرد این معماری در تشخیص شکستگی‌های مچ دست کودکان بر روی مجموعه‌داده GRAZPEDWRI-DXدر منابع بررسی‌شده یافت نشد. بنابراین، هنوز مشخص نیست که آیا مزایای گزارش‌شده برای یولو ۲۶ در حوزه تصاویر طبیعی، در تصاویر رادیوگرافی پزشکی نیز قابل دستیابی است یا خیر.

بر این اساس، خلأ پژوهشی این مطالعه صرفاً استفاده از یک نسخه جدید از خانواده یولو نیست، بلکه نبود ارزیابی مستقل و مستند از قابلیت‌های یولو ۲۶ در مسئله تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان بر روی یک مجموعه‌داده استاندارد پزشکی است. علاوه بر این، بررسی پژوهش‌های پیشین نشان داد که تفاوت در پروتکل‌های آموزش و ارزیابی، مقایسه منصفانه عملکرد مدل‌های مختلف را دشوار ساخته است. ازاین‌رو، انجام یک ارزیابی نظام‌مند با استفاده از پروتکل مشخص، معیارهای استاندارد و مجموعه‌داده مرجع می‌تواند شواهد قابل اتکاتری درباره توانایی‌های این معماری ارائه کند.

بر همین اساس، پژوهش حاضر با بهره‌گیری از مجموعه‌داده GRAZPEDWRI-DX، عملکرد مدل یولو ۲۶ را در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان مورد ارزیابی قرار می‌دهد. هدف این پژوهش، بررسی تجربی قابلیت‌ها و محدودیت‌های این معماری در یک کاربرد پزشکی واقعی و تحلیل عملکرد آن بر اساس معیارهای استاندارد آشکارسازی اشیا است. نتایج این ارزیابی می‌تواند علاوه بر روشن کردن ظرفیت یولو ۲۶ در این حوزه، مبنایی برای توسعه پژوهش‌های آینده و بهبود سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی فراهم سازد.

بر مبنای مباحث مطرح‌شده در این فصل، در فصل سوم روش اجرای پژوهش تشریح می‌شود. در آن فصل، نحوه آماده‌سازی مجموعه‌داده، پیش‌پردازش تصاویر، انتخاب و پیکربندی مدل یولو ۲۶ ، تنظیمات آموزش، پروتکل ارزیابی و محیط اجرای آزمایش‌ها به‌صورت دقیق معرفی خواهند شد تا امکان بازتولید و ارزیابی نتایج پژوهش فراهم شود.

فصل سوم روش اجرای پژوهش

۳-۱ مقدمه

پس از بررسی مبانی نظری، تحلیل پژوهش‌های پیشین و تبیین خلأ پژوهشی در فصل دوم، در این فصل روش اجرای پژوهش به‌صورت نظام‌مند تشریح می‌شود. هدف اصلی این فصل، ارائه فرآیند کامل انجام پژوهش به‌گونه‌ای است که تمامی مراحل از آماده‌سازی داده‌ها تا آموزش، ارزیابی و تحلیل عملکرد مدل، به‌صورت شفاف، مستند و قابل بازتولید بیان شوند. در پژوهش‌های مبتنی بر یادگیری عمیق، اعتبار نتایج تنها به انتخاب یک مدل مناسب وابسته نیست، بلکه طراحی صحیح پروتکل آزمایش، کنترل متغیرهای مؤثر، استفاده از معیارهای استاندارد ارزیابی و امکان بازتولید نتایج نیز از عوامل تعیین‌کننده در اعتبار علمی پژوهش محسوب می‌شوند.

بر همین اساس، در پژوهش حاضر تلاش شده است تمامی مراحل اجرای آزمایش بر پایه یک پروتکل از پیش تعیین‌شده و ثابت طراحی شوند تا امکان مقایسه منصفانه میان مدل‌های مورد بررسی فراهم شود. در این پروتکل، متغیرهایی نظیر مجموعه‌داده، نحوه تقسیم داده‌ها، اندازه تصاویر ورودی، مقدار بذر تصادفی[۲۴] معیار انتخاب مدل، سیاست افزایش داده‌ها و سایر تنظیمات مؤثر در فرآیند آموزش کنترل شده‌اند تا تفاوت نتایج تنها ناشی از ویژگی‌های مدل‌های مورد مقایسه باشد و نه تغییر شرایط اجرای آزمایش. این رویکرد، علاوه بر افزایش اعتبار نتایج، امکان بازتولید آزمایش‌ها توسط سایر پژوهشگران را نیز فراهم می‌کند.

به‌منظور مقایسه منصفانه عملکرد مدل‌ها، معماری ، یولو ۹و یولو ۲۶ در شرایط یکسان مورد بررسی قرار گرفته‌اند. تمامی مدل‌ها با استفاده از مجموعه‌داده یکسان، تقسیم داده ثابت، اندازه تصویر ورودی یکسان، مقدار سید مشخص و معیار ارزیابی مشترک آموزش و ارزیابی می‌شوند. علاوه بر این، تمامی تنظیمات آزمایش، مشخصات محیط اجرا و پارامترهای مورد استفاده ثبت شده‌اند تا فرآیند پژوهش از نظر علمی شفاف و قابل بازتولید باشد.

بر اساس ساختار ارائه‌شده، در ادامه این فصل ابتدا طراحی کلی پژوهش و پروتکل اجرای آزمایش معرفی می‌شود. سپس مجموعه‌داده مورد استفاده، مراحل آماده‌سازی و پیش‌پردازش داده‌ها، مدل‌های مورد بررسی، تنظیمات آموزش، محیط اجرای آزمایش و پروتکل ارزیابی تشریح خواهند شد. در پایان نیز مراحل اجرای آزمایش به‌صورت یک فرآیند منسجم ارائه می‌شود تا زمینه لازم برای ارائه نتایج و تحلیل آن‌ها در فصل چهارم فراهم گردد.

۳-۲ طراحی پژوهش و پروتکل اجرای آزمایش

پژوهش حاضر از نوع کاربردی و تجربی است و با رویکرد کمی انجام شده است. هدف اصلی پژوهش، ارزیابی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان و مقایسه آن با مدل‌های مرجع منتخب در یک محیط آزمایشی کنترل‌شده

[۲۴] Seed

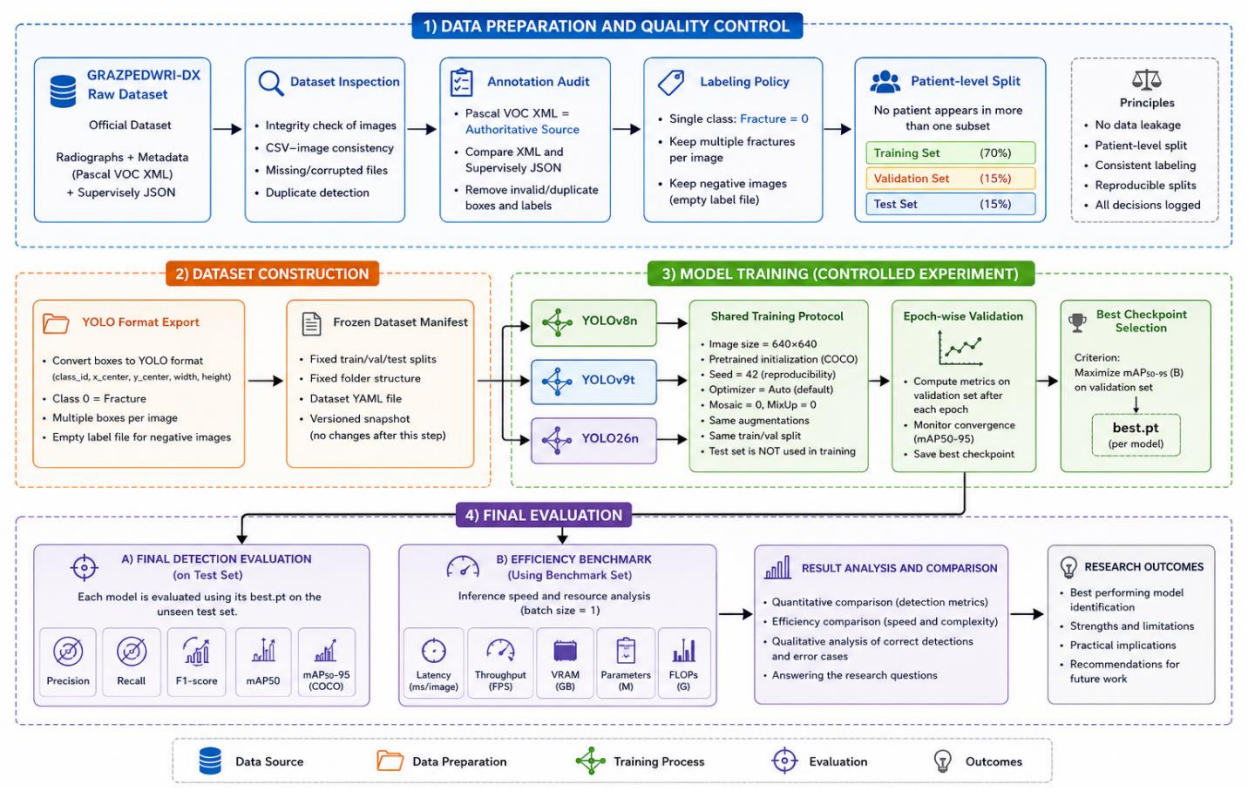
است. به همین منظور، فرآیند اجرای پژوهش بر اساس یک پروتکل آزمایش از پیش تعیین‌شده طراحی شده است تا تمامی مراحل آموزش و ارزیابی تحت شرایط یکسان انجام شوند و نتایج حاصل از مدل‌ها قابل مقایسه و قابل بازتولید باشند.

در طراحی این پروتکل، اصل اساسی بر کنترل متغیرهای مؤثر بر فرآیند آموزش استوار بوده است. به همین منظور، مجموعه‌داده، نحوه تقسیم داده‌ها، اندازه تصاویر ورودی، مقدار بذر تصادفی، سیاست افزایش داده‌ها، معیار انتخاب مدل و سایر پارامترهای مشترک برای تمامی مدل‌ها ثابت نگه داشته شدند و تنها معماری مدل به‌عنوان متغیر مستقل پژوهش تغییر می‌کند. این رویکرد موجب می‌شود اختلاف عملکرد مشاهده‌شده میان مدل‌ها ناشی از ویژگی‌های معماری آن‌ها باشد و نه تفاوت در شرایط اجرای آزمایش.

به‌منظور جلوگیری از ایجاد سوگیری در ارزیابی، تقسیم مجموعه‌داده در سطح بیمار انجام شده است؛ به‌گونه‌ای که تصاویر مربوط به یک بیمار تنها در یکی از مجموعه‌های آموزش، اعتبارسنجی یا آزمون قرار می‌گیرند. علاوه بر این، مجموعه آزمون در طول مراحل آموزش و انتخاب مدل کاملاً کنار گذاشته شده و تنها پس از پایان فرآیند آموزش و اعتبارسنجی برای ارزیابی نهایی مورد استفاده قرار می‌گیرد. این سیاست از نشت اطلاعات میان مجموعه‌های داده جلوگیری کرده و امکان ارزیابی واقع‌بینانه‌تر توانایی تعمیم مدل را فراهم می‌کند.

در این پژوهش، سه مدل YOLOv۸n، یولو ۹ tو یولو ۲۶ nبه‌عنوان مدل‌های مورد مقایسه انتخاب شده‌اند. تمامی مدل‌ها با استفاده از مجموعه‌داده یکسان، تقسیم داده ثابت و معیار انتخاب مشترک آموزش داده می‌شوند. همچنین، معیار انتخاب بهترین مدل در مرحله آموزش، $mAP_{۵۰-۹۵}$بر روی مجموعه اعتبارسنجی است و این معیار برای همه مدل‌ها بدون تغییر مورد استفاده قرار می‌گیرد. بدین ترتیب، مقایسه عملکرد مدل‌ها بر پایه یک چارچوب ارزیابی واحد انجام می‌شود.

از دیگر اصول طراحی این پژوهش، قابلیت بازتولید است. تمامی تنظیمات آزمایش، نسخه نرم‌افزارها، ساختار مجموعه‌داده، پارامترهای آموزش و شرایط اجرای آزمایش پیش از شروع فرآیند آموزش ثبت شده‌اند و در طول اجرای پژوهش بدون تغییر باقی می‌مانند. ثبت این اطلاعات امکان تکرار آزمایش‌ها توسط سایر پژوهشگران و بررسی مستقل نتایج را فراهم می‌سازد و اعتبار علمی پژوهش را افزایش می‌دهد.



شکل ۱-۳. نمای کلی فرآیند اجرای پژوهش

شکل ۱-۳ نمایی کلی از فرآیند اجرای پژوهش را نشان می‌دهد. همان‌گونه که مشاهده می‌شود، پژوهش از آماده‌سازی مجموعه‌داده آغاز شده، پس از آموزش و اعتبارسنجی مدل‌ها، ارزیابی نهایی بر روی مجموعه آزمون انجام می‌شود و در پایان، نتایج بر اساس معیارهای استاندارد آشکارسازی اشیا تحلیل و مقایسه می‌شوند.

در ادامه این فصل، اجزای تشکیل‌دهنده این پروتکل شامل مجموعه‌داده مورد استفاده، مراحل آماده‌سازی داده‌ها، مدل‌های مورد بررسی، تنظیمات آموزش، محیط اجرای آزمایش و روش ارزیابی عملکرد به‌صورت تفصیلی تشریح می‌شوند.

۳-۳ مجموعه‌داده مورد استفاده

با توجه به بررسی انجام‌شده در فصل دوم، مجموعه‌داده GRAZPEDWRI-DX به‌عنوان مجموعه‌داده مرجع این پژوهش انتخاب شد. این مجموعه‌داده به دلیل دسترس‌پذیری عمومی، کیفیت مناسب برچسب‌گذاری و استفاده گسترده در مطالعات مرتبط، بستر مناسبی برای آموزش و ارزیابی مدل‌های آشکارسازی اشیا در تشخیص شکستگی‌های مچ دست کودکان فراهم می‌کند. در این پژوهش، از نسخه رسمی این مجموعه‌داده استفاده شده و تمامی مراحل آماده‌سازی و اجرای آزمایش‌ها بر پایه همین داده‌ها انجام گرفته است.

مجموعه‌داده مورد استفاده شامل تصاویر رادیوگرافی و فایل‌های برچسب متناظر با هر تصویر است که بر اساس قالب مورد نیاز مدل‌های خانواده یولو سازمان‌دهی شده‌اند. به‌منظور سهولت در اجرای فرآیند آموزش و ارزیابی، داده‌ها پس از آماده‌سازی در سه مجموعه

آموزش[۲۵]، اعتبارسنجی[۲۶] و آزمون[۲۷] سازمان‌دهی شدند. هر یک از این مجموعه‌ها تنها در مرحله متناظر با خود مورد استفاده قرار گرفتند؛ به‌گونه‌ای که داده‌های آموزش صرفاً برای یادگیری مدل، داده‌های اعتبارسنجی برای پایش فرآیند آموزش و انتخاب بهترین مدل و داده‌های آزمون تنها برای ارزیابی نهایی عملکرد استفاده شدند.

به‌منظور حفظ یکنواختی فرآیند ارزیابی، ساختار تقسیم داده‌ها در تمامی آزمایش‌های انجام‌شده ثابت نگه داشته شد و هیچ تغییری در ترکیب مجموعه‌های آموزش، اعتبارسنجی و آزمون در طول پژوهش اعمال نشد. همچنین، مجموعه آزمون تا پایان فرآیند آموزش و انتخاب مدل کاملاً کنار گذاشته شد و تنها پس از اتمام آموزش برای ارزیابی نهایی عملکرد مورد استفاده قرار گرفت. این رویکرد موجب می‌شود نتایج گزارش‌شده تصویری واقع‌بینانه از توانایی تعمیم مدل به داده‌های دیده‌نشده ارائه کنند. جدول ۱-۳ نحوه تقسیم مجموعه‌داده مورد استفاده در پژوهش را نشان می‌دهد.

مجموعه	توضیح	هدف استفاده
آموزش	استفاده در فرآیند آموزش	یادگیری پارامترهای مدل
اعتبارسنجی	استفاده در طول آموزش بدون به‌روزرسانی نهایی مدل	پایش آموزش و انتخاب بهترین مدل
آزمون	استفاده تنها پس از پایان فرآیند آموزش	ارزیابی نهایی عملکرد

جدول ۱-۳. ساختار تقسیم مجموعه‌داده مورد استفاده

پس از سازمان‌دهی مجموعه‌داده و انجام فرآیند تقسیم، مرحله آماده‌سازی و پیش‌پردازش داده‌ها آغاز شد. این مرحله شامل تبدیل داده‌ها به قالب مورد نیاز مدل، آماده‌سازی تصاویر و اعمال عملیات لازم پیش از آموزش است که در بخش بعد به‌صورت تفصیلی تشریح می‌شود.

۳-۴ آماده‌سازی و پیش‌پردازش داده‌ها

به‌منظور استفاده مؤثر از مجموعه‌داده در فرآیند آموزش مدل، داده‌های اولیه پیش از آغاز آموزش آماده‌سازی و سازمان‌دهی شدند. هدف از این مرحله، ایجاد یک ساختار یکنواخت برای داده‌ها، اطمینان از صحت برچسب‌ها و فراهم کردن شرایط مناسب برای آموزش و ارزیابی مدل‌ها بود. تمامی عملیات آماده‌سازی بر اساس پروتکل تعریف‌شده پژوهش انجام شد تا در تمامی آزمایش‌ها از یک پایپلاین یکسان استفاده شود و قابلیت بازتولید نتایج حفظ گردد.

۳-۴-۱ آماده‌سازی اولیه داده‌ها

[۲۵] **Training**
[۲۶] **Validation**
[۲۷] **Test**

پس از دریافت نسخه رسمی مجموعه‌داده، ساختار فایل‌ها و برچسب‌ها بررسی شد و داده‌ها مطابق قالب مورد نیاز چارچوب آموزش مدل‌های خانواده یولو سازمان‌دهی شدند. تصاویر و فایل‌های برچسب متناظر در پوشه‌های مجزای آموزش، اعتبارسنجی و آزمون قرار گرفتند تا فرآیند آموزش بدون تغییر در ساختار داده‌ها انجام شود. همچنین، هماهنگی میان تصاویر و فایل‌های برچسب کنترل شد تا از وجود نمونه‌های ناقص یا ناسازگار جلوگیری شود. در این مرحله هیچ تغییری در محتوای تصاویر یا برچسب‌های مرجع اعمال نشد و تنها عملیات سازمان‌دهی و آماده‌سازی ساختار داده‌ها برای ورود به پایپلاین آموزش انجام گرفت.

۳-۴-۲ پیش‌پردازش تصاویر

به‌منظور ایجاد یکنواختی در داده‌های ورودی، تصاویر در هنگام ورود به پایپلاین آموزش با اندازه ورودی تعیین‌شده برای مدل پردازش شدند. این فرآیند موجب می‌شود تمامی تصاویر با ابعاد یکسان وارد شبکه شوند و شرایط آموزش برای تمام نمونه‌ها یکنواخت باشد. عملیات پایه مورد نیاز برای آماده‌سازی تصاویر، از جمله تبدیل قالب داده‌ها و نرمال‌سازی مقادیر پیکسل، توسط چارچوب آموزش مورد استفاده انجام شد و مرحله مستقلی برای این منظور خارج از پایپلاین آموزش تعریف نشد. به همین دلیل، تمامی مدل‌های مورد بررسی دقیقاً از یک فرآیند پیش‌پردازش استفاده کردند و تفاوتی در نحوه آماده‌سازی داده‌ها میان آن‌ها وجود نداشت.

۳-۴-۳ افزایش داده‌ها[۲۸]

برای افزایش تنوع داده‌های آموزشی و کاهش احتمال بیش‌برازش مدل، از قابلیت‌های افزایش داده موجود در چارچوب آموزش استفاده شد. عملیات افزایش داده تنها بر روی مجموعه آموزش اعمال گردید و مجموعه‌های اعتبارسنجی و آزمون بدون هیچ‌گونه تغییر مورد استفاده قرار گرفتند تا ارزیابی مدل بر اساس داده‌های واقعی انجام شود.

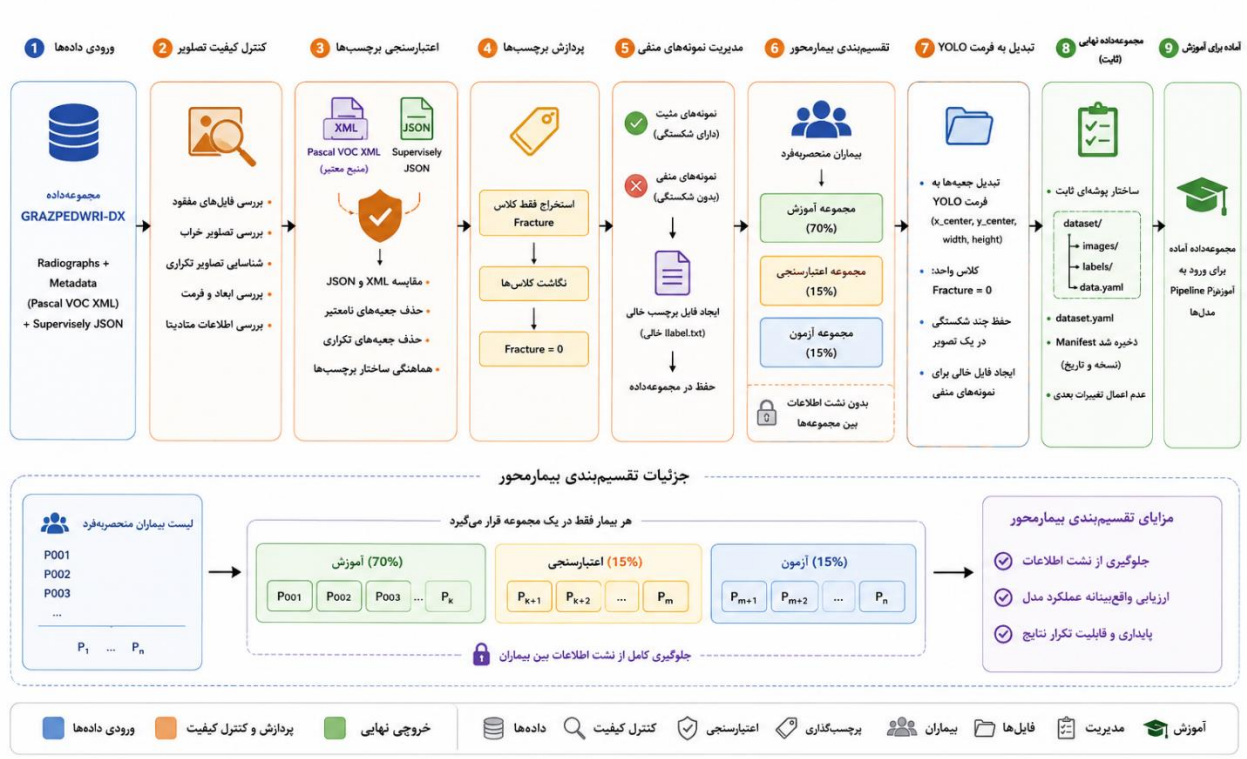
تمامی تنظیمات مربوط به افزایش داده پیش از آغاز آموزش تعیین و در طول اجرای آزمایش بدون تغییر حفظ شدند. همچنین، این تنظیمات برای همه مدل‌های مورد بررسی یکسان بودند تا شرایط مقایسه کاملاً منصفانه باقی بماند و تفاوت عملکرد مدل‌ها تنها ناشی از ویژگی‌های معماری آن‌ها باشد.

۳-۴-۴ جمع‌بندی فرآیند آماده‌سازی داده‌ها

فرآیند آماده‌سازی داده‌ها در این پژوهش با هدف ایجاد یک پایپلاین یکنواخت، قابل بازتولید و بدون سوگیری طراحی شد. سازمان‌دهی استاندارد داده‌ها، استفاده از پیش‌پردازش یکسان برای تمامی مدل‌ها و اعمال افزایش داده تنها بر روی مجموعه آموزش، موجب شد تمامی مدل‌ها تحت شرایط یکسان آموزش ببینند و نتایج حاصل از آن‌ها قابل مقایسه باشند.

شکل ۲-۳ نمای کلی پایپلاین آماده‌سازی داده‌ها را نشان می‌دهد.

[۲۸] **Data Augmentation**



شکل ۲-۳. نمای کلی Pipeline آماده‌سازی داده‌ها

۳-۵ مدل‌های مورد بررسی

پس از آماده‌سازی مجموعه‌داده، مرحله انتخاب مدل‌های مورد استفاده در آزمایش آغاز شد. هدف این پژوهش، ارزیابی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان و مقایسه آن با دو مدل مرجع از نسل‌های پیشین خانواده یولو است. برای آنکه نتایج حاصل از این مقایسه معتبر و قابل استناد باشند، مدل‌های مورد بررسی بر اساس یک چارچوب یکسان آموزش و ارزیابی شدند و تمامی شرایط آزمایش، به‌جز معماری مدل، ثابت نگه داشته شد.

۳-۵-۱ معیار انتخاب مدل‌ها

انتخاب مدل‌های مورد بررسی بر اساس دو معیار اصلی انجام شد. نخست، استفاده از مدل‌هایی که در مطالعات پیشین عملکرد مناسبی در مسئله آشکارسازی اشیا و به‌ویژه تصاویر پزشکی نشان داده‌اند و دوم، امکان مقایسه روند تکامل معماری‌های خانواده یولو با حفظ شرایط یکسان آموزش و ارزیابی. بر این اساس، سه مدل YOLOv۸n، یولو ۹ و یولو ۲۶ به‌عنوان مدل‌های مورد مطالعه انتخاب شدند. مدل یولو ۸ به‌عنوان یکی از مدل‌های مرجع و پرکاربرد در پژوهش‌های اخیر، نقش خط پایه را در این پژوهش ایفا می‌کند. یولو ۹ نیز به دلیل بهره‌گیری از بهبودهای معماری نسبت به نسل‌های قبل، به‌عنوان نماینده نسل بعدی خانواده یولو در مقایسه حضور دارد. در نهایت، یولو ۲۶ nبه‌عنوان جدیدترین معماری مورد بررسی و مدل اصلی پژوهش انتخاب شده است.

۳-۵-۲ مدل اصلی پژوهش

تمرکز اصلی این پژوهش بر ارزیابی مدل یولو ۲۶ nاست. این مدل بدون اعمال تغییر در ساختار معماری اصلی و مطابق با پیاده‌سازی رسمی آن مورد استفاده قرار گرفته است. بدین ترتیب، هدف پژوهش حاضر ارائه یک معماری جدید یا اصلاح ساختار داخلی مدل نیست، بلکه ارزیابی تجربی عملکرد نسخه استاندارد یولو ۲۶ در مسئله تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان است.

انتخاب نسخه Nano (n)نیز با هدف ایجاد تعادل میان سرعت پردازش، مصرف منابع محاسباتی و دقت مدل انجام شده است. این نسخه علاوه بر نیاز کمتر به حافظه و توان پردازشی، امکان مقایسه منصفانه با نسخه‌های هم‌رده مدل‌های مرجع را نیز فراهم می‌کند.

۳-۵-۳ چارچوب مقایسه مدل‌ها

به‌منظور آنکه نتایج حاصل از مقایسه مدل‌ها تنها تحت تأثیر تفاوت‌های معماری قرار گیرد، تمامی مدل‌های مورد بررسی با مجموعه‌داده یکسان، تقسیم داده ثابت، تنظیمات آموزشی مشترک و معیارهای ارزیابی یکسان آموزش و آزمون شدند. در نتیجه، هیچ تفاوتی در فرآیند آماده‌سازی داده‌ها، نحوه آموزش یا روش ارزیابی میان مدل‌ها وجود نداشت و تنها متغیر مستقل پژوهش، معماری مدل بود.

جدول ۲-۳ مدل‌های مورد استفاده در این پژوهش و نقش هر یک را نشان می‌دهد.

جدول ۲-۳. مدل‌های مورد استفاده در پژوهش

مدل	هدف استفاده	نقش در پژوهش
YOLOv۸n	ایجاد خط پایه برای مقایسه	مدل مرجع (Baseline)
یولو t۹	بررسی روند تکامل معماری‌های خانوادهYOLO	مدل مقایسه‌ای
یولو ۲۶ n	ارزیابی عملکرد در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان	مدل اصلی پژوهش

در این پژوهش، هیچ تغییری در معماری داخلی مدل‌های مورد بررسی اعمال نشد و تمامی آزمایش‌ها بر پایه پیاده‌سازی استاندارد هر مدل انجام شدند. این رویکرد موجب می‌شود نتایج به‌دست‌آمده بیانگر عملکرد واقعی هر معماری در شرایط یکسان آزمایش باشند و امکان مقایسه مستقیم میان مدل‌ها فراهم شود. در بخش بعد، تنظیمات آموزش، پارامترهای مشترک مورد استفاده در فرآیند یادگیری و محیط اجرای آزمایش‌ها به‌صورت دقیق معرفی می‌شوند.

۳-۶ محیط اجرای آزمایش و تنظیمات آموزش

به‌منظور تضمین قابلیت بازتولید نتایج و فراهم‌سازی امکان تکرار آزمایش‌ها توسط سایر پژوهشگران، تمامی مراحل آموزش، اعتبارسنجی و ارزیابی مدل‌های مورد بررسی در یک محیط سخت‌افزاری و نرم‌افزاری مشخص انجام شد. همچنین، تمامی مدل‌ها با استفاده از مجموعه‌داده یکسان، تنظیمات آموزشی مشترک و پروتکل ارزیابی مشابه آموزش داده شدند تا شرایط مقایسه‌ای یکسان برای آن‌ها فراهم

شود. بدین ترتیب، تفاوت‌های مشاهده‌شده در عملکرد مدل‌ها را می‌توان عمدتاً ناشی از تفاوت معماری آن‌ها دانست و اثر عوامل جانبی ناشی از محیط اجرا یا تنظیمات آموزش به حداقل رسید.

۳-۶-۱ محیط اجرای آزمایش

تمامی آزمایش‌های این پژوهش بر روی یک ایستگاه کاری مجهز به پردازنده گرافیکی NVIDIA GeForce RTX ۴۰۹۰ با ۲۴ گیگابایت حافظه اختصاصی انجام شد. فرآیند آموزش مدل‌ها با استفاده از چارچوب Ultralytics مبتنی بر PyTorch و با بهره‌گیری از قابلیت پردازش موازی CUDA اجرا گردید. همچنین، در طول انجام پژوهش نسخه سیستم‌عامل، زبان برنامه‌نویسی، کتابخانه‌های اصلی و سایر ابزارهای مورد استفاده بدون تغییر حفظ شدند تا از ایجاد اختلاف در نتایج ناشی از تفاوت محیط اجرایی جلوگیری شود. مشخصات کامل محیط اجرای آزمایش در جدول ۳-۳ ارائه شده است.

مؤلفه	مشخصات
سیستم‌عامل	Ubuntu Linux
زبان برنامه‌نویسی	Python ۳,۱۲,۱۳
چارچوب یادگیری عمیق	PyTorch ۲,۷,۱
کتابخانه آموزش مدل	Ultralytics ۸,٤,۱۰۵
شتاب‌دهنده محاسباتی	CUDA ۱۲,۸
پردازنده گرافیکی	NVIDIA GeForce RTX ٤۰۹۰ (۲٤ GB)
حافظه اصلی	۹۴ GB RAM

جدول ۳-۳. مشخصات محیط اجرای آزمایش

۳-۶-۲ تنظیمات آموزش

به‌منظور حفظ یکنواختی شرایط آزمایش و امکان مقایسه مستقیم عملکرد مدل‌ها، تمامی پارامترهای آموزشی پیش از آغاز فرآیند آموزش تعیین شده و در تمامی آزمایش‌ها بدون تغییر مورد استفاده قرار گرفتند. در هر سه مدل، اندازه تصویر ورودی ۶۴۰×۶۴۰ پیکسل، تعداد دوره‌های آموزش ۱۰۰، اندازه دسته آموزشی ۶۴، نرخ یادگیری اولیه ۰.۰۱، مقدار Weight Decay برابر ۰.۰۰۰۵ و مقدار Seed برابر ۴۲ در نظر گرفته شد. همچنین، به‌منظور جلوگیری از ادامه آموزش در صورت عدم بهبود عملکرد مدل، از مکانیزم Early Stopping با مقدار Patience برابر ۲۰ استفاده شد.

فرآیند انتخاب بهترین مدل صرفاً بر اساس عملکرد آن بر روی مجموعه اعتبارسنجی انجام گرفت و مجموعه آزمون تا پایان مرحله آموزش مورد استفاده قرار نگرفت. بدین ترتیب، از تأثیر داده‌های آزمون بر فرآیند یادگیری جلوگیری شد و ارزیابی نهایی بر روی داده‌های کاملاً دیده‌نشده انجام پذیرفت.

علاوه بر این، تمامی مدل‌ها با استفاده از مجموعه‌داده یکسان، تقسیم‌بندی ثابت داده‌ها و تنظیمات آموزشی مشابه آموزش داده شدند و هیچ‌گونه تنظیم اختصاصی برای یک مدل خاص اعمال نشد. این رویکرد موجب شد شرایط آموزش برای تمامی مدل‌ها یکسان بوده و اختلاف نتایج حاصل را بتوان به تفاوت‌های معماری آن‌ها نسبت داد.

پارامتر	مقدار
اندازه تصویر ورودی	۶۴۰ × ۶۴۰ پیکسل
تعداد دوره‌های آموزش(Epoch)	۱۰۰
اندازه دسته آموزشی(Batch Size)	۶۴
نرخ یادگیری اولیه	۰.۰۱
Weight Decay	۰.۰۰۰۵
Optimizer	Auto
Seed	۴۲
Early Stopping (Patience)	۲۰

جدول ۴-۳. تنظیمات مشترک آموزش مدل‌ها

در ادامه، روش ارزیابی عملکرد مدل‌ها و معیارهای مورد استفاده برای تحلیل نتایج معرفی می‌شود.

۳-۷ روش ارزیابی عملکرد مدل‌ها

پس از پایان فرآیند آموزش، عملکرد مدل‌های مورد بررسی بر روی مجموعه آزمون ارزیابی شد. به‌منظور آنکه نتایج حاصل از آزمایش‌ها قابل مقایسه، قابل تفسیر و منطبق با استانداردهای رایج در حوزه آشکارسازی اشیا باشند، ارزیابی تمامی مدل‌ها با استفاده از مجموعه‌ای از معیارهای استاندارد و تحت یک پروتکل یکسان انجام شد. تمامی شاخص‌های ارزیابی از خروجی مدل‌ها بر روی داده‌های دیده‌نشده استخراج شدند و هیچ‌یک از داده‌های آزمون در فرآیند آموزش یا انتخاب مدل مورد استفاده قرار نگرفتند.

همان‌گونه که در فصل دوم بیان شد، برای ارزیابی عملکرد مدل‌ها از معیارهای دقت، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۵۰ درصد و میانگین دقت در بازه آستانه‌های هم‌پوشانی ۵۰ تا ۹۵ درصد استفاده شد. در این پژوهش، میانگین دقت در بازه آستانه‌های هم‌پوشانی ۵۰ تا ۹۵ درصد به‌عنوان شاخص اصلی مقایسه عملکرد مدل‌ها انتخاب شد؛ زیرا این معیار با ارزیابی عملکرد مدل در طیفی از آستانه‌های هم‌پوشانی، تصویر جامع‌تر و دقیق‌تری از توانایی آن در تشخیص و مکان‌یابی شکستگی ارائه می‌کند. سایر معیارها نیز به‌عنوان شاخص‌های تکمیلی برای تحلیل عملکرد مدل‌ها مورد استفاده قرار گرفتند.

برای حفظ یکنواختی فرآیند ارزیابی، تمامی مدل‌ها بر روی مجموعه آزمون یکسان و با تنظیمات استنتاج مشابه مورد ارزیابی قرار گرفتند. علاوه بر این، بهترین نسخه هر مدل بر اساس عملکرد آن بر روی مجموعه اعتبارسنجی انتخاب و همان نسخه برای ارزیابی نهایی مورد

استفاده قرار گرفت. بنابراین، تمامی مدل‌ها تحت شرایط یکسان ارزیابی شدند و اختلاف نتایج صرفاً بیانگر تفاوت عملکرد آن‌ها در مسئله تشخیص شکستگی است.

معیار	کاربرد در پژوهش
دقت مثبت	ارزیابی میزان صحت پیش‌بینی‌های مثبت مدل
فراخوانی	ارزیابی توانایی مدل در شناسایی شکستگی‌ها
mAP@۰٫۵	مقایسه عملکرد مدل در آستانه هم‌پوشانی ۰٫۵
mAP@۰٫۵:۰٫۹۵	معیار اصلی مقایسه و انتخاب عملکرد مدل‌ها

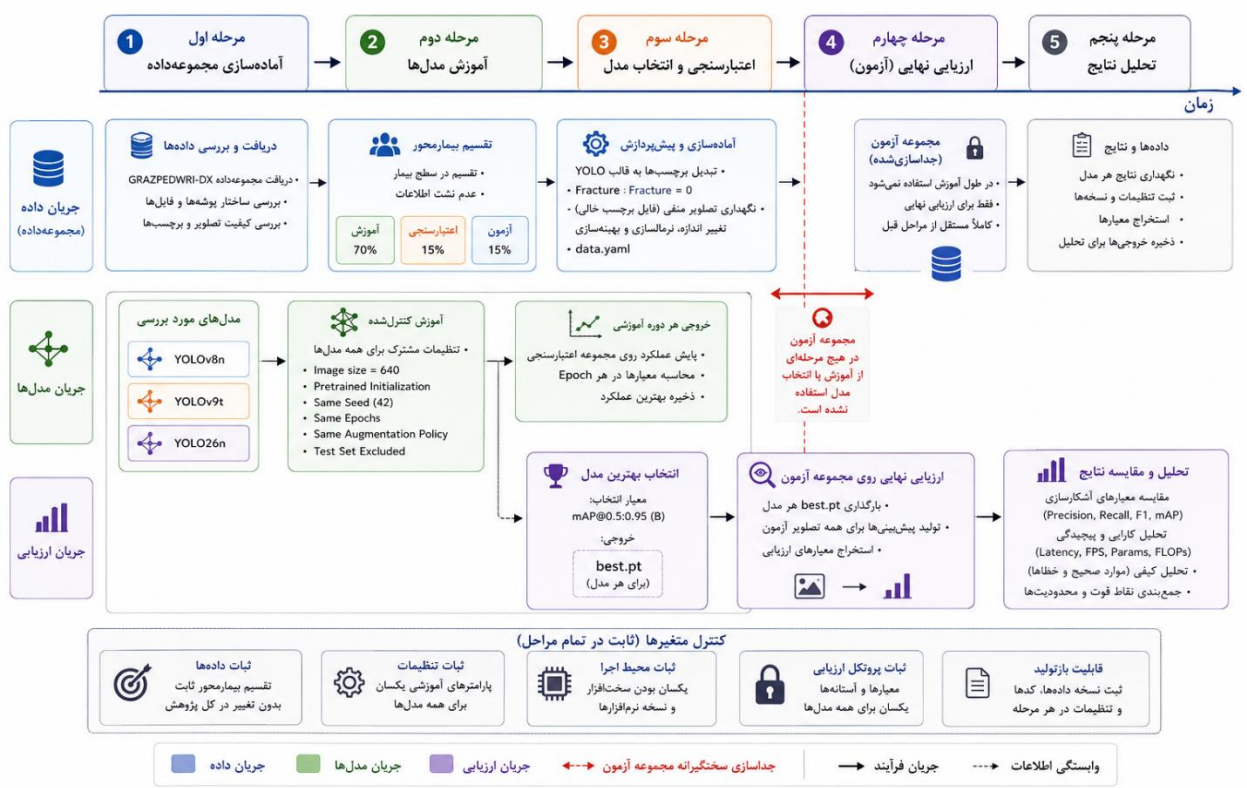
جدول ۵-۳. معیارهای ارزیابی مورد استفاده در پژوهش

تمامی معیارهای فوق با استفاده از خروجی استاندارد چارچوب Ultralytics محاسبه شدند و هیچ‌گونه معیار سفارشی یا تغییر یافته‌ای در فرآیند ارزیابی مورد استفاده قرار نگرفت. استفاده از معیارهای استاندارد، علاوه بر افزایش قابلیت اعتماد نتایج، امکان مقایسه مستقیم یافته‌های این پژوهش با مطالعات پیشین را نیز فراهم می‌کند.

بر اساس روش ارزیابی تشریح‌شده در این فصل، نتایج حاصل از آموزش و ارزیابی مدل‌های مورد بررسی در فصل چهارم ارائه و تحلیل خواهند شد.

۸-۳ مراحل اجرای آزمایش

به‌منظور ارزیابی عملکرد مدل‌های مورد بررسی، فرآیند اجرای پژوهش بر اساس پروتکل طراحی‌شده و در قالب مجموعه‌ای از مراحل متوالی انجام شد. در تمامی این مراحل، ترتیب اجرای عملیات ثابت نگه داشته شد و هر مرحله پس از تکمیل مرحله قبل آغاز گردید. این رویکرد ضمن حفظ انسجام فرآیند پژوهش، امکان بازتولید کامل آزمایش‌ها را نیز فراهم می‌سازد. شکل ۳-۳ نمای کلی مراحل اجرای پژوهش را نشان می‌دهد.



شکل ۳-۳. نمای کلی مراحل اجرای پژوهش

مرحله اول: آماده‌سازی مجموعه‌داده

در نخستین مرحله، مجموعه‌داده دریافت و پس از بررسی ساختار داده‌ها، بر اساس پروتکل پژوهش سازمان‌دهی شد. سپس تقسیم داده‌ها در سطح بیمار انجام گرفت و مجموعه‌های آموزش، اعتبارسنجی و آزمون ایجاد شدند. در ادامه، عملیات آماده‌سازی و پیش‌پردازش داده‌ها برای ورود به فرآیند آموزش انجام شد.

مرحله دوم: آموزش مدل‌ها

پس از آماده‌سازی داده‌ها، مدل‌های مورد بررسی با استفاده از مجموعه آموزش و بر اساس تنظیمات مشترک تعریف‌شده آموزش داده شدند. تمامی مدل‌ها تحت شرایط یکسان و با استفاده از پارامترهای آموزشی ثابت اجرا شدند تا اثر متغیرهای جانبی بر نتایج به حداقل برسد.

مرحله سوم: اعتبارسنجی و انتخاب مدل

در طول فرآیند آموزش، عملکرد مدل‌ها بر روی مجموعه اعتبارسنجی به‌صورت پیوسته پایش شد. پس از پایان آموزش، بهترین نسخه هر مدل بر اساس معیار انتخاب از پیش تعیین‌شده انتخاب و برای ارزیابی نهایی ذخیره شد.

مرحله چهارم: ارزیابی نهایی

در این مرحله، بهترین نسخه هر مدل بر روی مجموعه آزمون که در طول آموزش مورد استفاده قرار نگرفته بود، اجرا شد. خروجی مدل‌ها برای تمامی نمونه‌های آزمون تولید و نتایج مورد نیاز برای ارزیابی عملکرد استخراج گردید.

مرحله پنجم: تحلیل نتایج

در پایان، معیارهای ارزیابی محاسبه و نتایج حاصل از مدل‌های مختلف با یکدیگر مقایسه شدند. علاوه بر تحلیل کمی شاخص‌های عملکرد، نتایج به‌منظور بررسی نقاط قوت و محدودیت‌های هر مدل مورد تحلیل قرار گرفتند. یافته‌های این مرحله در فصل چهارم به‌صورت تفصیلی ارائه خواهند شد.

۹-۳ جمع‌بندی فصل

در این فصل، روش اجرای پژوهش و چارچوب آزمایش به‌صورت نظام‌مند تشریح شد. ابتدا طراحی پژوهش و اصول حاکم بر پروتکل اجرای آزمایش معرفی گردید و سپس نحوه استفاده از مجموعه‌داده، مراحل آماده‌سازی و پیش‌پردازش داده‌ها، مدل‌های مورد بررسی، تنظیمات آموزش، محیط اجرای آزمایش و روش ارزیابی عملکرد مدل‌ها ارائه شد. همچنین، فرآیند اجرای آزمایش از آماده‌سازی داده‌ها تا ارزیابی نهایی به‌صورت گام‌به‌گام تبیین گردید تا امکان بازتولید کامل پژوهش فراهم شود.

در طراحی این پژوهش تلاش شد تمامی عوامل مؤثر بر اعتبار نتایج کنترل شوند. استفاده از تقسیم داده در سطح بیمار، جداسازی کامل مجموعه آزمون از فرآیند آموزش، به‌کارگیری یک پروتکل ثابت برای تمامی مدل‌ها و ثبت دقیق تنظیمات اجرای آزمایش، از جمله اقداماتی هستند که با هدف افزایش قابلیت اعتماد، کاهش سوگیری و فراهم کردن امکان مقایسه منصفانه میان مدل‌های مورد بررسی انجام شده‌اند. بدین ترتیب، چارچوب روش‌شناسی پژوهش به‌گونه‌ای طراحی شده است که نتایج حاصل از آزمایش‌ها تا حد امکان ناشی از تفاوت‌های واقعی میان مدل‌ها باشند و نه تفاوت در شرایط اجرای آزمایش.

بر اساس روش تشریح‌شده در این فصل، در فصل چهارم نتایج حاصل از آموزش و ارزیابی مدل‌های مورد بررسی ارائه و تحلیل می‌شوند. در آن فصل، عملکرد مدل‌ها بر اساس معیارهای استاندارد ارزیابی شده و ضمن مقایسه نتایج، نقاط قوت و محدودیت‌های هر مدل مورد بررسی قرار خواهد گرفت تا تصویری روشن از توانایی مدل در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان ارائه شود.

فصل چهارم نتایج

۴-۱ مقدمه

در فصل‌های پیشین، مسئله پژوهش، مبانی نظری، مطالعات مرتبط و روش انجام پژوهش به‌صورت جامع تشریح شد. همچنین مجموعه‌داده مورد استفاده، مدل‌های منتخب، معیارهای ارزیابی و تنظیمات کلی اجرای آزمایش‌ها معرفی شدند. پس از آماده‌سازی داده‌ها و آموزش مدل‌های مورد بررسی، گام بعدی ارزیابی عملکرد آن‌ها و تحلیل نتایج حاصل از آزمایش‌ها است.

هدف این فصل، ارائه و تحلیل نتایج حاصل از ارزیابی مدل‌های مورد بررسی در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی است. بدین منظور، عملکرد مدل‌ها بر اساس معیارهای استاندارد ارزیابی آشکارسازی اشیا، شامل دقت مثبت ، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ ، مورد بررسی قرار می‌گیرد. علاوه بر شاخص‌های کمی، روند آموزش مدل‌ها، کیفیت مکان‌یابی شکستگی‌ها و نمونه‌هایی از خروجی‌های حاصل نیز به‌منظور ارائه تصویری جامع از عملکرد مدل‌ها تحلیل خواهند شد.

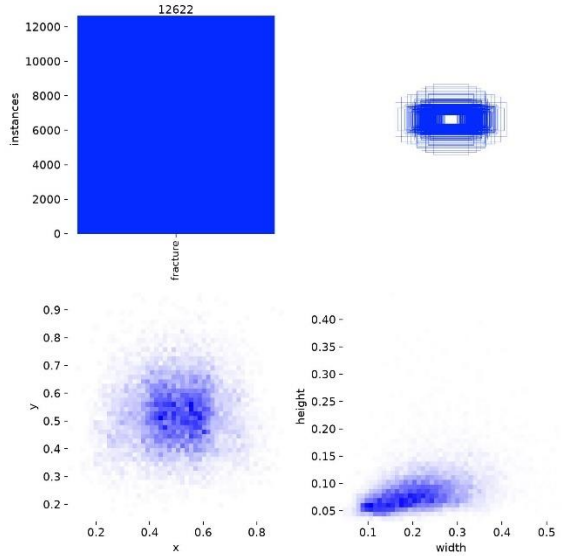
از آنجا که تمامی مدل‌ها با استفاده از مجموعه‌داده یکسان، شرایط آموزشی مشابه و معیارهای ارزیابی یکسان مورد بررسی قرار گرفته‌اند، نتایج ارائه‌شده امکان مقایسه مستقیم عملکرد آن‌ها را فراهم می‌سازد. این مقایسه علاوه بر مشخص کردن میزان دقت و کارایی هر مدل، نقاط قوت و محدودیت‌های آن‌ها را نیز آشکار کرده و مبنایی برای انتخاب مناسب‌ترین مدل در کاربرد مورد مطالعه فراهم می‌آورد.

بر این اساس، در ادامه این فصل ابتدا روند آموزش و همگرایی مدل‌ها مورد بررسی قرار می‌گیرد. سپس عملکرد نهایی آن‌ها بر اساس معیارهای کمی ارزیابی مقایسه می‌شود. در ادامه، نمونه‌هایی از خروجی‌های مدل‌ها و مهم‌ترین خطاهای مشاهده‌شده تحلیل شده و در پایان، جمع‌بندی نتایج حاصل از آزمایش‌ها ارائه خواهد شد.

۴-۲ مشخصات داده‌های مورد استفاده در ارزیابی

مجموعه‌داده مورد استفاده در این پژوهش شامل تصاویر رادیوگرافی مچ دست کودکان و نوجوانان بوده و برای آموزش و ارزیابی سامانه‌های هوشمند تشخیص شکستگی توسعه یافته است. تصاویر این مجموعه‌داده از معاینات واقعی بیماران جمع‌آوری شده‌اند و نواحی آسیب‌دیده توسط متخصصان پزشکی در قالب جعبه‌های محدودکننده نشانه‌گذاری شده‌اند. وجود برچسب‌های مکانی، این مجموعه‌داده را برای آموزش مدل‌های آشکارسازی اشیا و ارزیابی هم‌زمان توانایی تشخیص و مکان‌یابی شکستگی مناسب ساخته است.

در پژوهش حاضر، مسئله آشکارسازی به‌صورت تک‌کلاسه تعریف شده و تمامی نواحی آسیب‌دیده با برچسب شکستگی مشخص شده‌اند. بر اساس توزیع برچسب‌های مجموعه آموزشی، تعداد ۱۲٫۶۲۲ نمونه شکستگی در داده‌های مورد استفاده برای آموزش مدل‌ها وجود دارد. نمودار توزیع برچسب‌ها علاوه بر تعداد نمونه‌ها، موقعیت مرکز جعبه‌های محدودکننده و توزیع عرض و ارتفاع آن‌ها را نیز نمایش می‌دهد.



شکل ۱-۴. توزیع تعداد، موقعیت و ابعاد جعبه‌های محدودکننده مربوط به شکستگی‌ها در مجموعه آموزشی

مطابق شکل ۱-۴، مرکز بخش عمده جعبه‌های محدودکننده در نواحی میانی تصویر قرار گرفته است. این الگو با ماهیت تصاویر رادیوگرافی مچ دست سازگار است؛ زیرا ناحیه مورد بررسی معمولاً در مرکز میدان تصویربرداری قرار می‌گیرد. با این حال، پراکندگی مرکز جعبه‌ها در بخش قابل‌توجهی از سطح تصویر نشان می‌دهد که موقعیت شکستگی‌ها کاملاً ثابت نیست و مدل باید توانایی تشخیص ضایعه را در موقعیت‌های متفاوت داشته باشد.

بررسی توزیع عرض و ارتفاع جعبه‌های محدودکننده نیز نشان می‌دهد که بیشتر شکستگی‌ها بخش نسبتاً کوچکی از تصویر را اشغال می‌کنند. عرض نرمال‌شده بیشتر جعبه‌ها تقریباً در بازه ۰٫۱ تا ۰٫۳ و ارتفاع آن‌ها عمدتاً در بازه ۰٫۰۵ تا ۰٫۱۵ قرار دارد. این موضوع نشان می‌دهد که بخش زیادی از نمونه‌های مورد بررسی را می‌توان در گروه اشیای کوچک یا متوسط دسته‌بندی کرد. تشخیص چنین نواحی‌ای نسبت به اشیای بزرگ دشوارتر است؛ زیرا اطلاعات تصویری مرتبط با شکستگی در تعداد محدودتری از پیکسل‌ها نمایش داده می‌شود و ممکن است تحت تأثیر تغییر اندازه تصویر، کیفیت رادیوگرافی و هم‌پوشانی ساختارهای استخوانی قرار گیرد.

به‌منظور فراهم کردن شرایط مقایسه منصفانه، هر سه مدل YOLOv۸، یولو ۹ و یولو ۲۶ با استفاده از یک مجموعه‌داده و تقسیم‌بندی یکسان آموزش و اعتبارسنجی شدند. یکسان بودن فایل تعریف مجموعه‌داده و مقدار هش ثبت‌شده برای آن در اجرای هر سه مدل نشان

می‌دهد که هیچ تغییری در نمونه‌های آموزشی و اعتبارسنجی میان آزمایش‌ها ایجاد نشده است. بنابراین، اختلاف مشاهده‌شده در عملکرد مدل‌ها را می‌توان عمدتاً به تفاوت معماری و رفتار آموزشی آن‌ها نسبت داد.

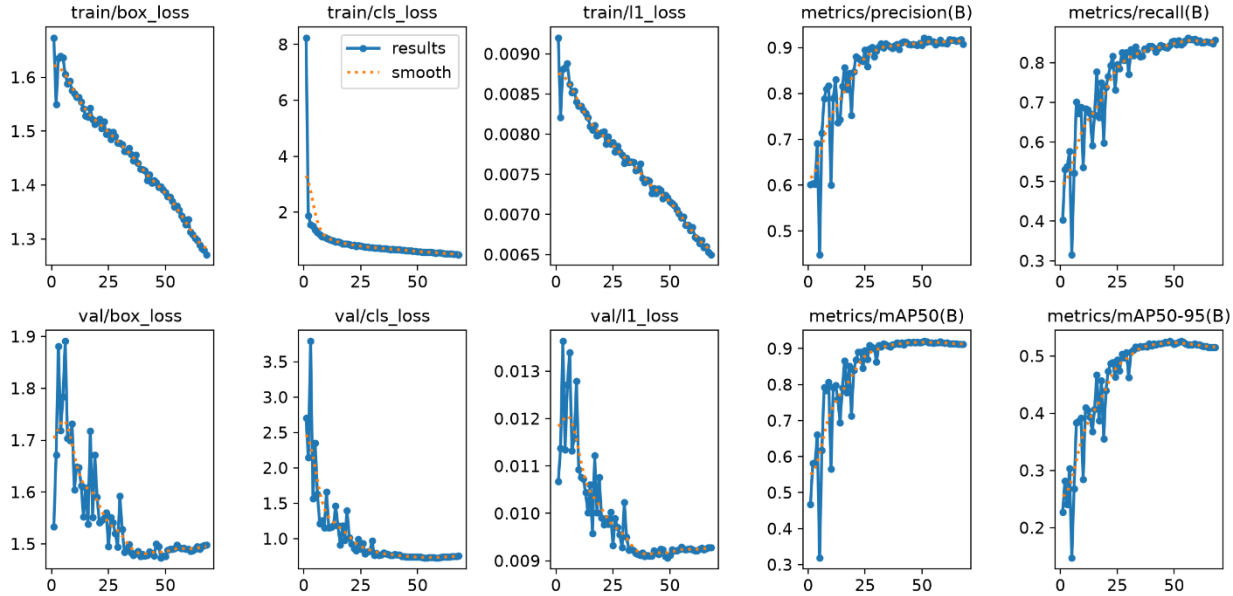
همچنین، در هر سه آزمایش، افزایش داده‌های مبتنی بر تغییر رنگ، ترکیب تصاویر و موزاییک غیرفعال بوده است. این تصمیم موجب شد مدل‌ها تحت یک پروتکل آموزشی کنترل‌شده و بدون تأثیر متفاوت روش‌های افزایش داده ارزیابی شوند. در نتیجه، مقایسه نتایج به‌صورت مستقیم‌تر انجام می‌شود و اثر تفاوت‌های معماری مدل‌ها با وضوح بیشتری قابل بررسی خواهد بود.

۴-۳ نتایج آموزش مدل یولو ۹

به‌منظور ارزیابی عملکرد مدل یولو ۹ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان، این مدل مطابق تنظیمات آموزشی ارائه‌شده در فصل سوم آموزش داده شد و عملکرد آن بر روی داده‌های اعتبارسنجی و آزمون مورد بررسی قرار گرفت. در طول فرآیند آموزش، تغییرات توابع خطا و شاخص‌های ارزیابی در پایان هر دوره ثبت شد تا روند یادگیری و میزان همگرایی مدل بررسی شود. پس از پایان آموزش نیز خروجی‌های استاندارد چارچوب آموزش، شامل منحنی‌های ارزیابی و ماتریس‌های درهم‌ریختگی، استخراج و برای تحلیل عملکرد مدل مورد استفاده قرار گرفت.

در این بخش، ابتدا روند آموزش و همگرایی مدل بررسی می‌شود. سپس عملکرد نهایی آن بر اساس منحنی‌های ارزیابی، ماتریس‌های درهم‌ریختگی و شاخص‌های کمی تحلیل خواهد شد.

۴-۳-۱ روند آموزش و همگرایی مدل



شکل ۱-۴: روند تغییر معیارهای آموزش و اعتبارسنجی مدل یولو ۹ در طول فرآیند آموزش

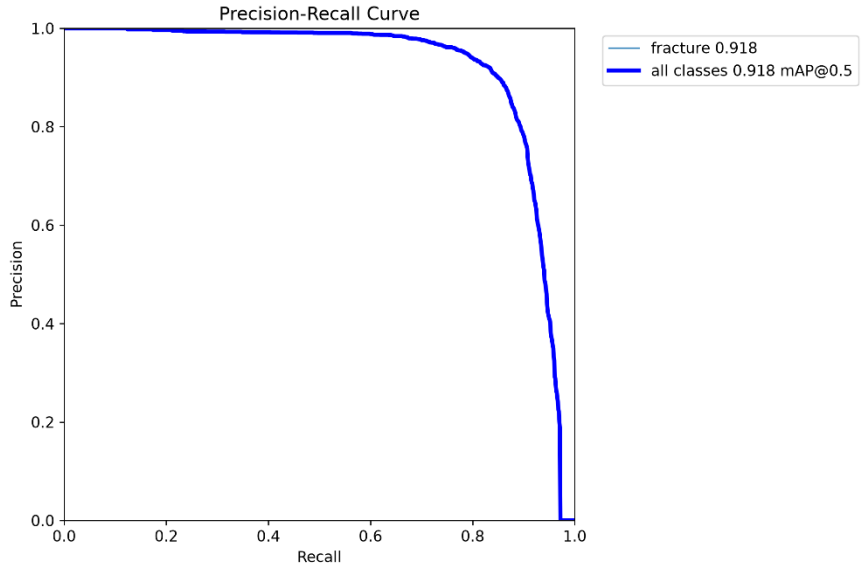
روند تغییرات معیارهای آموزش و اعتبارسنجی مدل در طول فرآیند آموزش در شکل ۱-۴ ارائه شده است. این شکل تغییرات تابع خطای مکان‌یابی، تابع خطای طبقه‌بندی، تابع خطای توزیع کانونی، همچنین تغییرات دقت مثبت، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ را در طول ۱۰۰ دوره آموزش نشان می‌دهد.

همان‌گونه که در شکل ۱-۴ مشاهده می‌شود، مقادیر توابع خطا در هر دو بخش آموزش و اعتبارسنجی در ابتدای فرآیند آموزش با سرعت بیشتری کاهش یافته و سپس به‌تدریج به وضعیت پایدار نزدیک شده‌اند. کاهش یکنواخت توابع خطا بیانگر آن است که مدل به‌تدریج ویژگی‌های مرتبط با شکستگی را فرا گرفته و توانایی مکان‌یابی نواحی آسیب‌دیده را بهبود داده است.

از سوی دیگر، روند افزایشی دقت مثبت، فراخوانی و میانگین دقت در طول آموزش نشان می‌دهد که با ادامه فرآیند یادگیری، عملکرد مدل به‌صورت تدریجی بهبود یافته است. بیشترین میزان تغییرات در دوره‌های ابتدایی آموزش مشاهده می‌شود و پس از آن، با نزدیک شدن مدل به نقطه همگرایی، سرعت تغییرات کاهش یافته و مقادیر شاخص‌های ارزیابی تقریباً ثابت می‌شوند. این رفتار بیانگر پایداری فرآیند آموزش و مناسب بودن تنظیمات انتخاب‌شده برای یادگیری مدل است.

همچنین، شباهت روند تغییرات معیارهای آموزش و اعتبارسنجی نشان می‌دهد که اختلاف محسوسی میان عملکرد مدل بر روی این دو مجموعه وجود ندارد و نشانه بارزی از بیش‌برازش در فرآیند آموزش مشاهده نمی‌شود. بنابراین، می‌توان نتیجه گرفت که مدل توانسته است علاوه بر یادگیری مناسب داده‌های آموزشی، قابلیت تعمیم مطلوبی نیز بر روی داده‌های دیده‌نشده به دست آورد.

۲-۳-۴ منحنی دقت مثبت—فراخوانی

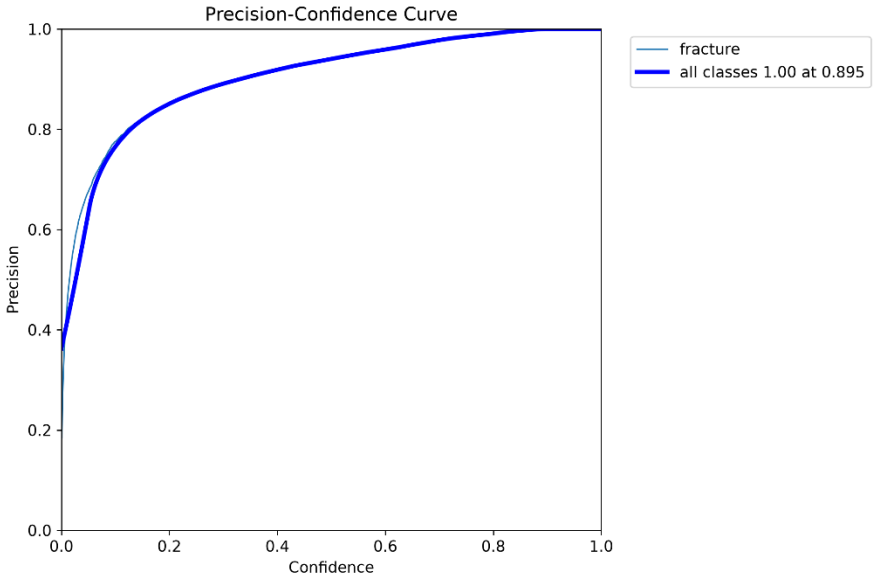


به‌منظور بررسی هم‌زمان رابطه میان دقت مثبت و فراخوانی، منحنی دقت مثبت—فراخوانی مدل در شکل ۲-۴ ارائه شده است. این منحنی یکی از مهم‌ترین معیارهای ارزیابی مدل‌های آشکارسازی اشیا محسوب می‌شود و هرچه مساحت زیر آن بیشتر باشد، عملکرد کلی مدل مطلوب‌تر خواهد بود.

شکل ۲-۴: منحنی دقت مثبت—فراخوانی مدل یولو ۹

مطابق شکل ۲-۴، منحنی دقت مثبت—فراخوانی در بخش عمده‌ای از دامنه فراخوانی در مقادیر بالای دقت مثبت قرار دارد که نشان‌دهنده عملکرد مناسب مدل در تشخیص شکستگی‌ها است. تنها در مقادیر بسیار بالای فراخوانی، کاهش تدریجی دقت مثبت مشاهده می‌شود که در مدل‌های آشکارسازی اشیا رفتاری طبیعی محسوب می‌شود.

۳-۳-۴ منحنی تغییرات دقت مثبت بر حسب آستانه اطمینان

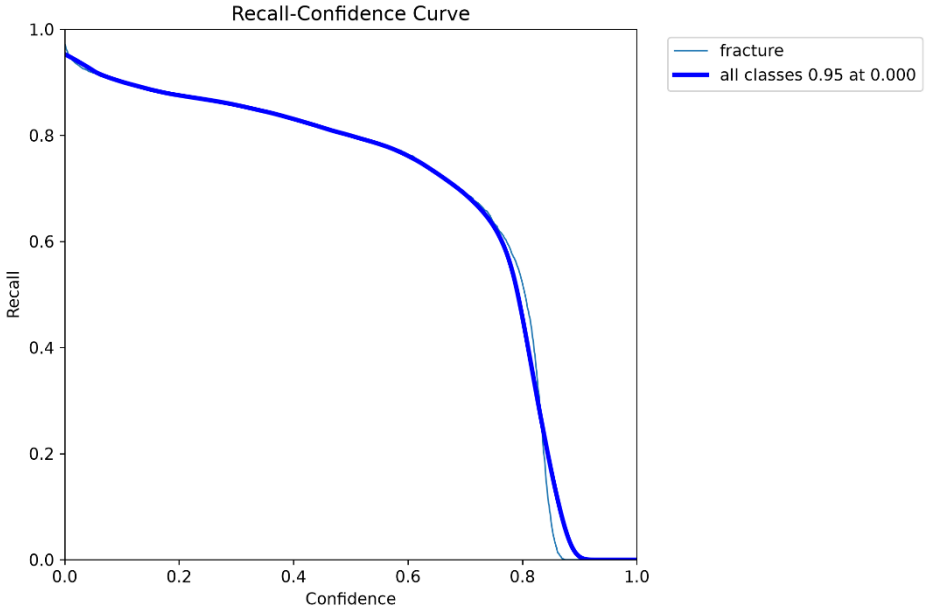


شکل ۳-۴ تغییرات دقت مثبت را در آستانه‌های مختلف اطمینان نمایش می‌دهد. این منحنی نشان می‌دهد که با افزایش آستانه اطمینان، پیش‌بینی‌های مدل با اطمینان بیشتری پذیرفته شده و در نتیجه، دقت مثبت افزایش می‌یابد.

شکل ۳-۴: تغییرات دقت مثبت بر حسب آستانه اطمینان در مدل یولو ۹

بر اساس شکل ۳-۴، با افزایش آستانه اطمینان، مقدار دقت مثبت نیز روندی افزایشی دارد و در آستانه‌های بالاتر به بیشترین مقدار خود نزدیک می‌شود. این رفتار نشان می‌دهد که پیش‌بینی‌های دارای اطمینان بیشتر، احتمال صحیح بودن بالاتری دارند.

۴-۳-۴ منحنی تغییرات فراخوانی بر حسب آستانه اطمینان



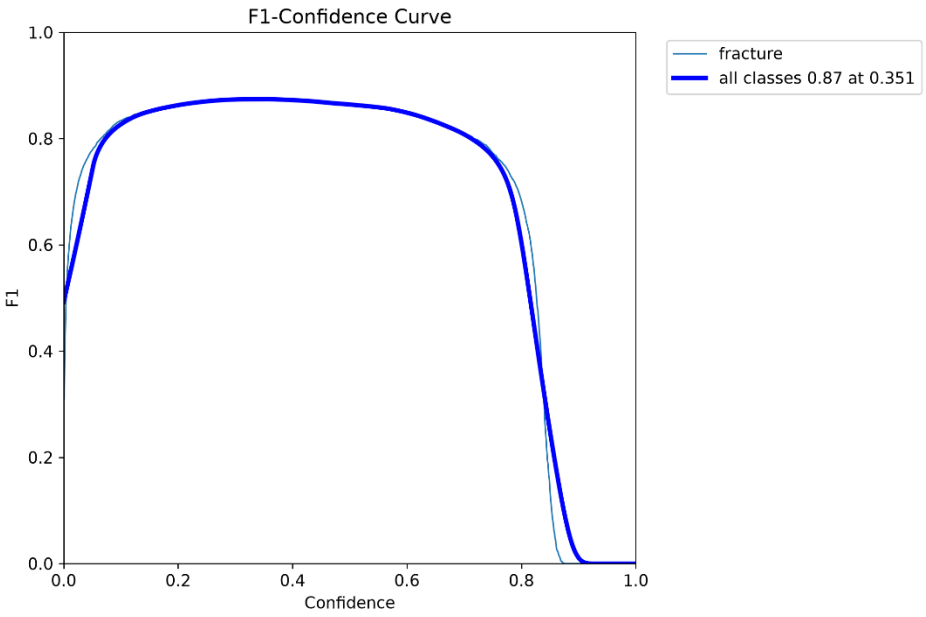
به‌منظور بررسی تأثیر آستانه اطمینان بر توانایی مدل در شناسایی شکستگی‌ها، منحنی تغییرات فراخوانی در شکل ۴-۴ ارائه شده است.

شکل ۴-۴: تغییرات فراخوانی بر حسب آستانه اطمینان در مدل یولو ۹

مطابق شکل ۴-۴، با افزایش آستانه اطمینان، مقدار فراخوانی به‌تدریج کاهش می‌یابد. علت این موضوع آن است که با سخت‌گیرانه‌تر شدن معیار پذیرش پیش‌بینی‌ها، بخشی از نمونه‌های صحیح نیز حذف می‌شوند. این رفتار نشان‌دهنده رابطه متقابل میان دقت مثبت و فراخوانی است و در مدل‌های آشکارسازی اشیا کاملاً طبیعی تلقی می‌شود.

۴-۳-۵ منحنی ضریب F۱ بر حسب آستانه اطمینان

ضریب F۱ یکی از مهم‌ترین شاخص‌های ارزیابی است که تعادل میان دقت مثبت و فراخوانی را نشان می‌دهد. شکل ۴-۵ تغییرات این شاخص را در آستانه‌های مختلف اطمینان نمایش می‌دهد.

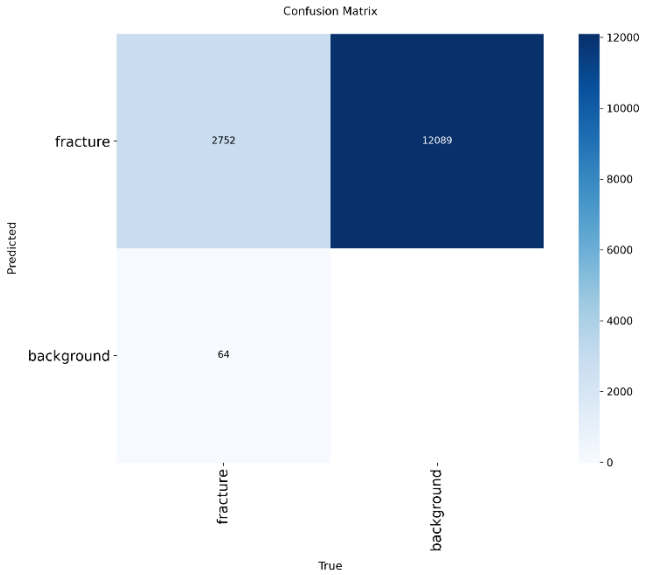


شکل ۵-۴: تغییرات ضریب $F^{۱}$ بر حسب آستانه اطمینان در مدل یولو ۹

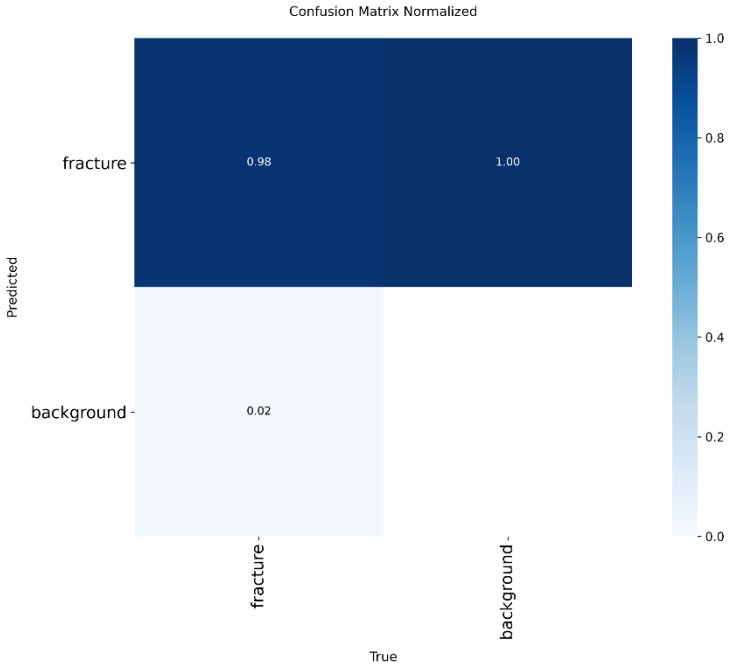
همان‌گونه که در شکل ۵-۴ مشاهده می‌شود، مقدار ضریب $F^{۱}$ در یک بازه مشخص از آستانه اطمینان به بیشترین مقدار خود می‌رسد که بیانگر ایجاد تعادل مناسب میان دقت مثبت و فراخوانی است. در مقادیر کمتر یا بیشتر از این بازه، کاهش ضریب $F^{۱}$ مشاهده می‌شود که ناشی از کاهش یکی از دو شاخص اصلی ارزیابی است.

۶-۳-۴ ماتریس درهم‌ریختگی

به‌منظور بررسی نحوه توزیع پیش‌بینی‌های صحیح و خطاهای مدل، ماتریس درهم‌ریختگی و نسخه نرمال‌شده آن استخراج شد. این ماتریس‌ها تعداد نمونه‌های تشخیص صحیح، نمونه‌های از دست‌رفته و پیش‌بینی‌های نادرست را نمایش می‌دهند.



شکل ۶-۴: ماتریس درهم‌ریختگی مدل یولو ۹



شکل ۷-۴: ماتریس درهم‌ریختگی نرمال‌شده مدل یولو ۹

بررسی شکل‌های ۶-۴ و ۷-۴ نشان می‌دهد که بخش عمده نمونه‌های شکستگی به‌درستی شناسایی شده‌اند و میزان خطا در تشخیص کلاس اصلی پایین است. همچنین، بخش قابل توجهی از خطاهای مشاهده‌شده مربوط به نواحی پس‌زمینه است که در مدل‌های آشکارسازی اشیا به دلیل نحوه تطبیق جعبه‌های پیشنهادی با نمونه‌های واقعی، رفتاری متداول محسوب می‌شود.

۷-۳-۴ جمع‌بندی عملکرد مدل

نتایج حاصل از آموزش و ارزیابی نشان می‌دهد که این مدل توانسته است روند یادگیری پایدار و همگرایی مناسبی را در طول آموزش تجربه کند. کاهش یکنواخت توابع خطا، بهبود تدریجی شاخص‌های ارزیابی و نتایج حاصل از منحنی‌های عملکرد، بیانگر توانایی مناسب مدل در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان است. جدول ۱-۴ خلاصه شاخص‌های نهایی عملکرد این مدل را بر روی مجموعه آزمون نشان می‌دهد.

شاخص ارزیابی	مقدار
دقت مثبت	۰.۹۱۶۹۹
فراخوانی	۰.۸۵۵۲۳
میانگین دقت در آستانه ۰٫۵	۰.۹۲۶۰۶
میانگین دقت در بازه ۰٫۵ تا ۰٫۹۵	۰.۵۳۵۶۲

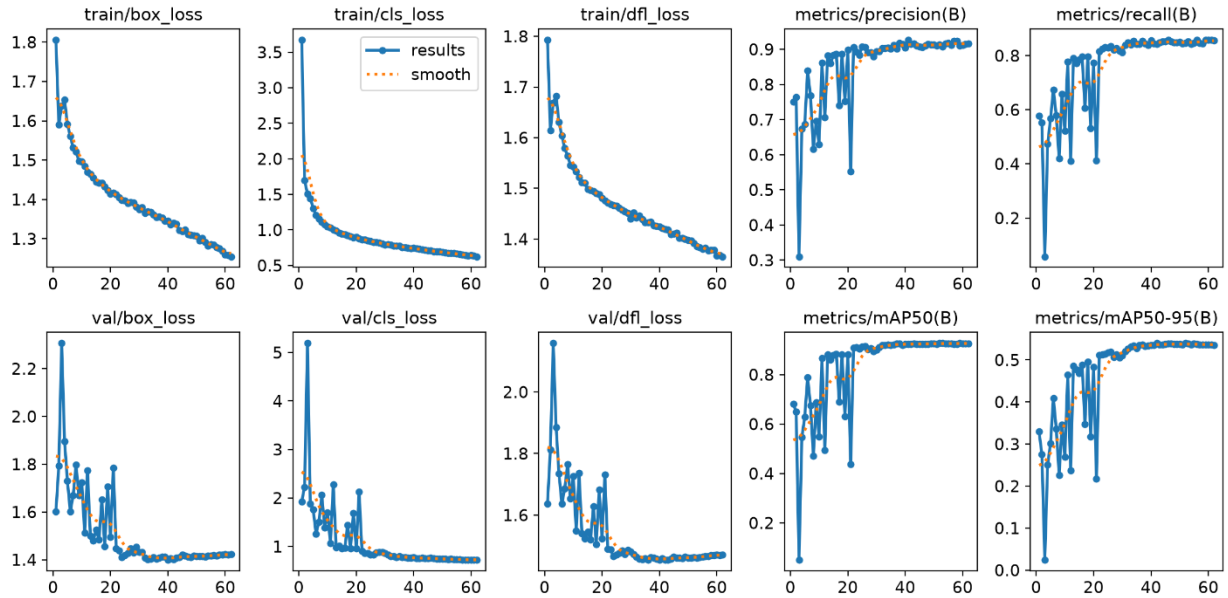
۴-۴ نتایج آموزش مدل یولو ۲۶

به‌منظور ارزیابی عملکرد مدل یولو ۲۶ در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان، این مدل مطابق تنظیمات آموزشی ارائه‌شده در فصل سوم آموزش داده شد و عملکرد آن بر روی مجموعه اعتبارسنجی و آزمون مورد بررسی قرار گرفت. در طول فرآیند آموزش، تغییرات توابع خطا و شاخص‌های ارزیابی در پایان هر دوره ثبت شد تا روند یادگیری و میزان همگرایی مدل بررسی شود. پس از پایان آموزش نیز خروجی‌های استاندارد چارچوب آموزش، شامل منحنی‌های ارزیابی و ماتریس‌های درهم‌ریختگی، استخراج و برای تحلیل عملکرد مدل مورد استفاده قرار گرفت.

در این بخش، ابتدا روند آموزش و همگرایی مدل بررسی شده و سپس عملکرد نهایی آن بر اساس منحنی‌های ارزیابی، ماتریس‌های درهم‌ریختگی و شاخص‌های کمی تحلیل می‌شود.

۱-۴-۴ روند آموزش و همگرایی مدل

روند تغییرات معیارهای آموزش و اعتبارسنجی مدل در طول فرآیند آموزش در شکل ۹-۴ ارائه شده است. این شکل تغییرات تابع خطای مکان‌یابی، تابع خطای طبقه‌بندی، تابع خطای توزیع کانونی، همچنین تغییرات دقت مثبت، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ را در طول فرآیند آموزش نمایش می‌دهد.

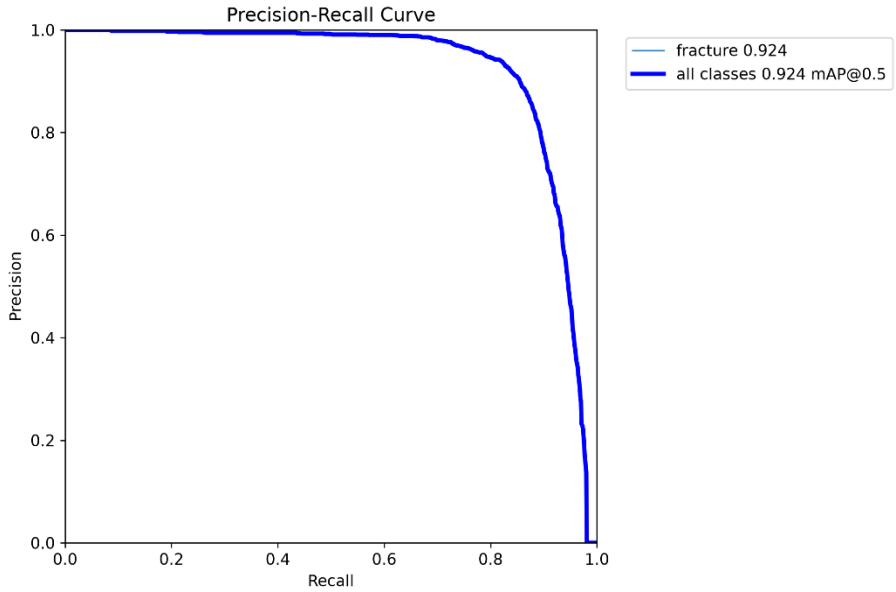


شکل ۹-۴: روند تغییر معیارهای آموزش و اعتبارسنجی مدل یولو ۲۶ در طول فرآیند آموزش

همان‌گونه که در شکل ۹-۴ مشاهده می‌شود، مقادیر توابع خطا در مراحل ابتدایی آموزش روندی کاهشی داشته و با ادامه فرآیند یادگیری به‌تدریج به حالت پایدار نزدیک شده‌اند. همچنین، شاخص‌های ارزیابی شامل دقت مثبت، فراخوانی و میانگین دقت نیز در طول آموزش روندی افزایشی داشته و در دوره‌های پایانی به وضعیت نسبتاً پایدار رسیده‌اند. این روند نشان‌دهنده همگرایی مناسب مدل و یادگیری تدریجی ویژگی‌های مرتبط با شکستگی‌های مچ دست است.

۲-۴-۴ منحنی دقت مثبت—فراخوانی

به‌منظور بررسی رابطه میان دقت مثبت و فراخوانی، منحنی دقت مثبت—فراخوانی مدل در شکل ۱۰-۴ ارائه شده است. این منحنی عملکرد کلی مدل را در آستانه‌های مختلف اطمینان نمایش داده و یکی از مهم‌ترین معیارهای ارزیابی مدل‌های آشکارسازی اشیا محسوب می‌شود.

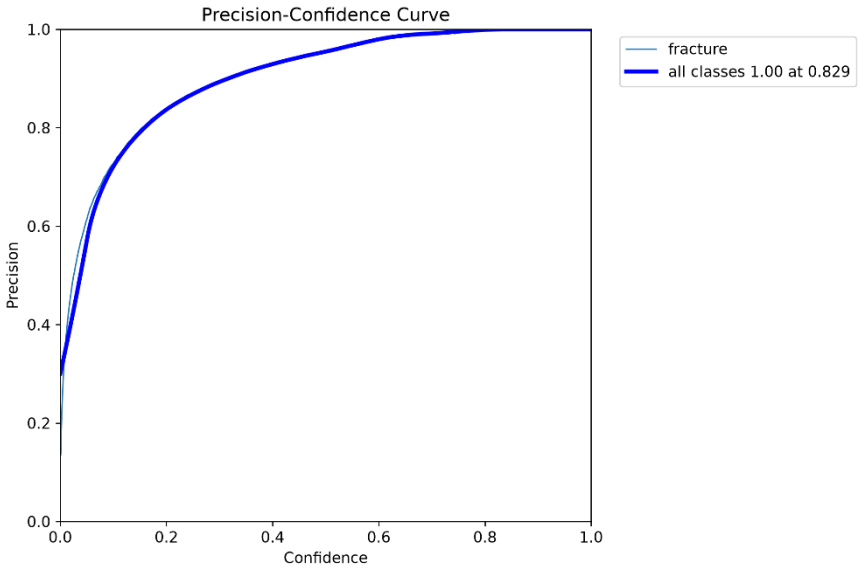


شکل ۱۰-۴: منحنی دقت مثبت—فراخوانی مدل یولو ۲۶

بررسی شکل ۱۰-۴ نشان می‌دهد که منحنی دقت مثبت—فراخوانی در بخش عمده‌ای از دامنه فراخوانی در مقادیر بالای دقت مثبت قرار دارد. با افزایش فراخوانی و نزدیک شدن به بیشینه آن، کاهش تدریجی دقت مثبت مشاهده می‌شود که رفتاری طبیعی در مدل‌های آشکارسازی اشیا است و بیانگر ایجاد تعادل میان این دو شاخص ارزیابی است.

۴-۴-۳ منحنی تغییرات دقت مثبت بر حسب آستانه اطمینان

شکل ۱۱-۴ تغییرات دقت مثبت را در آستانه‌های مختلف اطمینان نمایش می‌دهد و تأثیر تغییر مقدار آستانه اطمینان بر صحت پیش‌بینی‌های مدل را نشان می‌دهد.

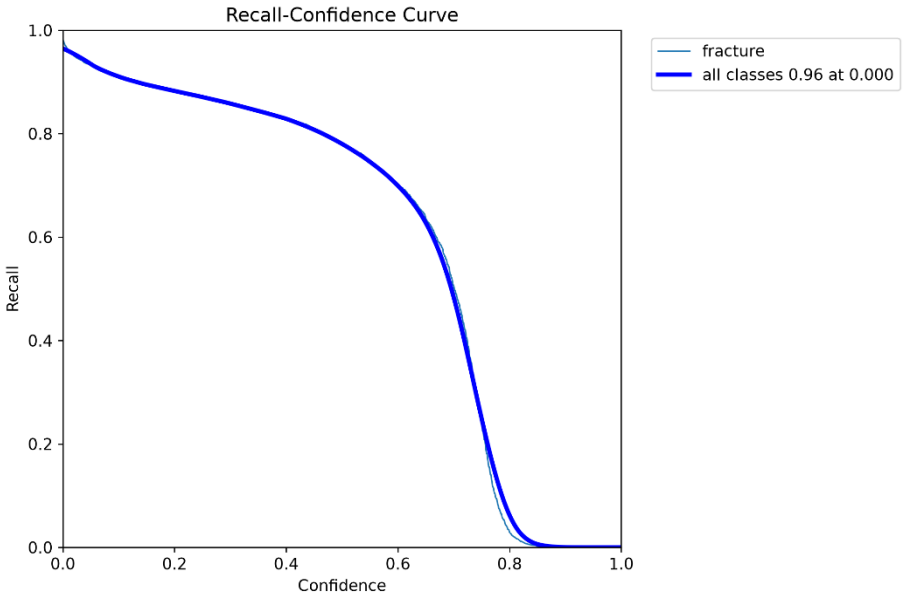


شکل ۱۱-۴: تغییرات دقت مثبت بر حسب آستانه اطمینان در مدل یولو ۲۶

مطابق شکل ۱۱-۴، با افزایش آستانه اطمینان، مقدار دقت مثبت نیز افزایش یافته و در آستانه‌های بالاتر به بیشترین مقدار خود نزدیک می‌شود. این رفتار نشان می‌دهد که پیش‌بینی‌های دارای اطمینان بیشتر، از صحت بالاتری برخوردار هستند.

۴-۴-۴ منحنی تغییرات فراخوانی بر حسب آستانه اطمینان

به‌منظور بررسی تأثیر آستانه اطمینان بر توانایی مدل در شناسایی شکستگی‌ها، منحنی تغییرات فراخوانی در شکل ۱۲-۴ ارائه شده است.

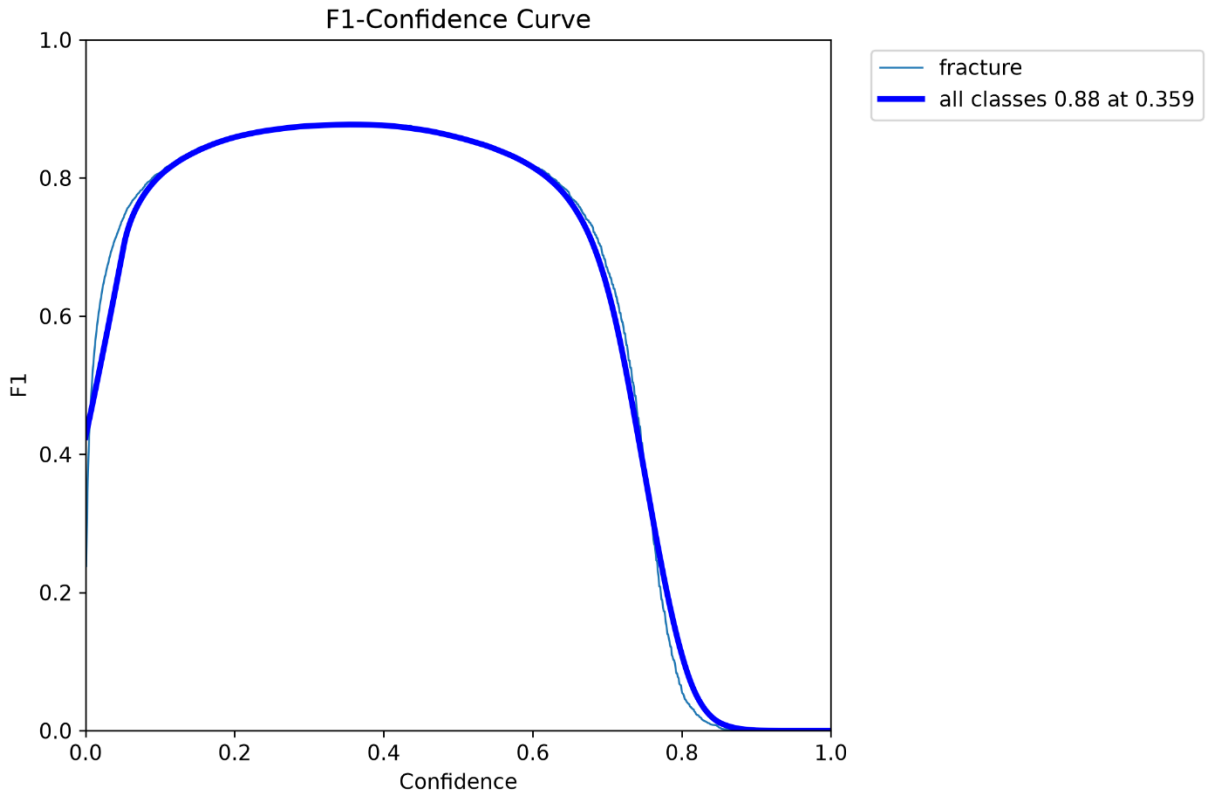


شکل ۱۲-۴: تغییرات فراخوانی بر حسب آستانه اطمینان در مدل یولو ۲۶

همان‌گونه که در شکل ۱۲-۴ مشاهده می‌شود، با افزایش آستانه اطمینان، مقدار فراخوانی به‌تدریج کاهش می‌یابد. این رفتار نشان می‌دهد که سخت‌گیرانه‌تر شدن معیار پذیرش پیش‌بینی‌ها، اگرچه موجب افزایش دقت مثبت می‌شود، اما بخشی از نمونه‌های صحیح نیز شناسایی نخواهند شد.

۵-۴-۴ منحنی ضریب F۱ بر حسب آستانه اطمینان

ضریب F۱ یکی از مهم‌ترین شاخص‌های ارزیابی عملکرد است که تعادل میان دقت مثبت و فراخوانی را نشان می‌دهد. تغییرات این شاخص در آستانه‌های مختلف اطمینان در شکل ۱۳-۴ ارائه شده است.

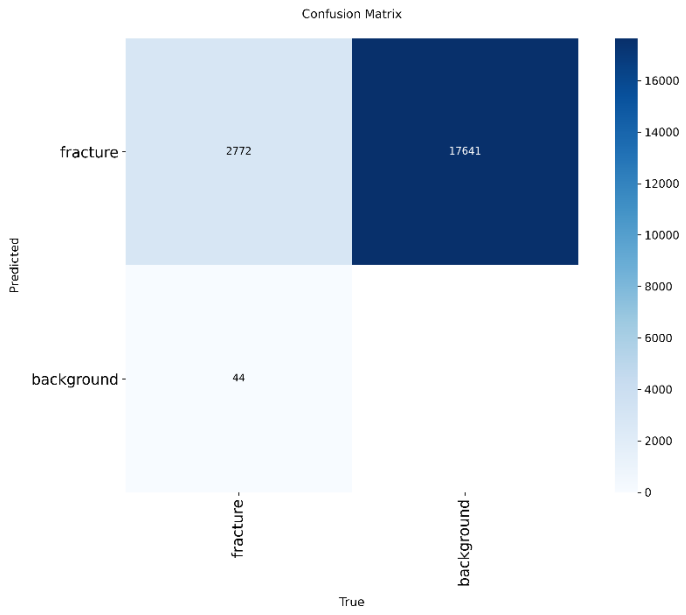


شکل ۱۳-۴:تغییرات ضریب F^1 بر حسب آستانه اطمینان در مدل یولو ۲۶

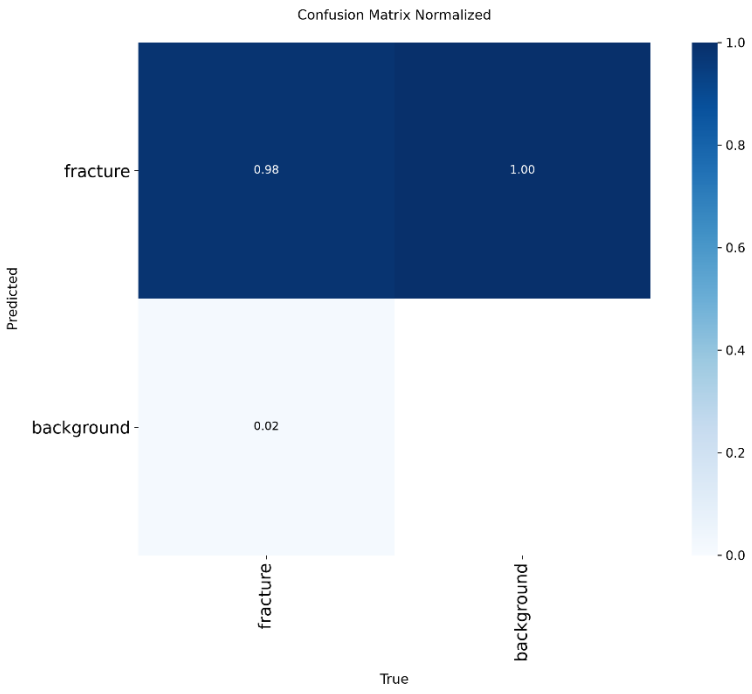
بررسی شکل ۱۳-۴ نشان می‌دهد که ضریب F^1 در یک بازه مشخص از آستانه اطمینان به بیشترین مقدار خود می‌رسد و سپس با افزایش بیشتر آستانه، روندی کاهشی پیدا می‌کند. این موضوع بیانگر وجود یک نقطه تعادل مناسب میان دقت مثبت و فراخوانی برای این مدل است.

۶-۴-۴ ماتریس درهم‌ریختگی

به‌منظور بررسی نحوه توزیع پیش‌بینی‌های صحیح و خطاهای مدل، ماتریس درهم‌ریختگی و نسخه نرمال‌شده آن استخراج شد. این ماتریس‌ها تعداد نمونه‌های تشخیص صحیح، نمونه‌های از دست‌رفته و پیش‌بینی‌های نادرست را نمایش می‌دهند.



شکل ۱۴-۴: ماتریس درهم‌ریختگی مدل یولو ۲۶



شکل ۱۵-۴: ماتریس درهم‌ریختگی نرمال‌شده مدل یولو ۲۶

بررسی شکل‌های ۱۴-۴ و ۱۵-۴ نشان می‌دهد که بخش عمده نمونه‌های شکستگی به‌درستی شناسایی شده‌اند و میزان خطا در تشخیص کلاس اصلی پایین است. همچنین، عمده پیش‌بینی‌های نادرست مربوط به نواحی پس‌زمینه است که در مدل‌های آشکارسازی اشیا، به دلیل نحوه تطبیق جعبه‌های پیشنهادی با نمونه‌های واقعی، رفتاری متداول محسوب می‌شود.

۷-۴-۴ جمع‌بندی عملکرد مدل

نتایج حاصل از آموزش و ارزیابی نشان می‌دهد که مدل یولو ۲۶ نیز از روند یادگیری پایدار و همگرایی مناسبی برخوردار بوده است. کاهش یکنواخت توابع خطا، بهبود تدریجی شاخص‌های ارزیابی و نتایج حاصل از منحنی‌های عملکرد، بیانگر توانایی مناسب این مدل در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان است. شاخص‌های نهایی عملکرد این مدل در جدول ۲-۴ ارائه شده‌اند و در بخش بعدی فصل، عملکرد آن به‌صورت مستقیم با مدل یولو ۹ مقایسه خواهد شد.

شاخص ارزیابی	مقدار
دقت مثبت	۰.۹۰۷۸۷
فراخوانی	۰.۸۵۷۳۷
میانگین دقت در آستانه ۰٫۵	۰.۹۱۲۰۵
میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵	۰.۵۱۵۴۳

جدول ۲-۴: شاخص‌های نهایی عملکرد مدل یولو ۲۶ بر روی مجموعه آزمون

۵-۴ مقایسه عملکرد مدل‌های مورد بررسی

پس از آموزش و ارزیابی مدل‌های مورد بررسی، عملکرد آن‌ها بر اساس شاخص‌های استاندارد ارزیابی شامل دقت مثبت، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ با یکدیگر مقایسه شد. از آنجا که تمامی مدل‌ها با استفاده از مجموعه‌داده یکسان، تنظیمات آموزشی مشابه و شرایط یکسان آموزش و ارزیابی شدند، اختلاف مشاهده‌شده در نتایج را می‌توان به تفاوت‌های معماری مدل‌ها نسبت داد. جدول ۳-۴ نتایج نهایی سه مدل مورد بررسی را بر اساس شاخص‌های ارزیابی نشان می‌دهد.

مدل	میانگین دقت در آستانه ۰٫۵	فراخوانی	دقت مثبت	میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵
YOLOv۸	۰.۹۲۲۴۲	۰.۸۵۲۶۳	۰.۹۱۰۹۵	۰.۵۲۶۸۵
یولو ۹	۰.۹۲۶۰۶	۰.۸۵۵۲۳	۰.۹۱۶۹۹	۰.۵۳۵۶۲
یولو ۲۶	۰.۹۱۲۰۵	۰.۸۵۷۳۷	۰.۹۰۷۸۷	۰.۵۱۵۴۳

بررسی نتایج جدول ۳-۴ نشان می‌دهد که هر سه مدل عملکرد قابل قبولی در تشخیص شکستگی‌های مچ دست کودکان داشته‌اند و اختلاف عملکرد آن‌ها نسبتاً محدود است. این موضوع نشان می‌دهد که هر سه معماری توانسته‌اند ویژگی‌های مرتبط با شکستگی را به‌خوبی فرا گیرند و بر روی مجموعه آزمون عملکرد مناسبی ارائه دهند.

از نظر دقت مثبت، مدل یولو ۹ با مقدار ۰.۹۱۶۹۹بهترین عملکرد را داشته و پس از آن مدل‌های YOLOv۸و یولو ۲۶ قرار گرفته‌اند. این موضوع بیانگر آن است که یولو ۹ در مقایسه با سایر مدل‌ها، پیش‌بینی‌های نادرست کمتری تولید کرده است.

در شاخص فراخوانی، مدل یولو ۲۶ با مقدار ۰.۸۵۷۳۷اندکی عملکرد بهتری نسبت به دو مدل دیگر داشته است. هرچند اختلاف این مقدار با مدل یولو ۹ بسیار ناچیز است، اما نشان می‌دهد یولو ۲۶ توانسته است تعداد بیشتری از نمونه‌های واقعی شکستگی را شناسایی کند.

همچنین، مقایسه میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ نشان می‌دهد که مدل یولو ۹ در هر دو شاخص بالاترین مقادیر را کسب کرده است. با توجه به اینکه معیار میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ به‌عنوان شاخص اصلی ارزیابی در این پژوهش انتخاب شده است، می‌توان نتیجه گرفت که مدل یولو ۹ در مجموع بهترین عملکرد را در میان مدل‌های مورد بررسی ارائه کرده است.

با وجود برتری مدل یولو ۹، اختلاف مقادیر شاخص‌های ارزیابی میان سه مدل نسبتاً محدود است. این موضوع بیانگر آن است که هر سه مدل از توانایی مناسبی در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان برخوردار هستند و انتخاب مدل نهایی می‌تواند علاوه بر دقت، با توجه به ملاحظاتی نظیر پیچیدگی معماری، هزینه محاسباتی و الزامات کاربردی نیز انجام شود.

۴-۶ جمع‌بندی فصل

در این فصل، نتایج حاصل از آموزش و ارزیابی مدل‌های مورد بررسی برای تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان ارائه و تحلیل شد. ابتدا عملکرد دو مدل یولو ۹ و یولو ۲۶ به‌صورت مستقل مورد بررسی قرار گرفت و روند آموزش، همگرایی مدل‌ها، منحنی‌های ارزیابی و ماتریس‌های درهم‌ریختگی هر یک تحلیل شد. نتایج به‌دست‌آمده نشان داد که هر دو مدل از روند یادگیری پایدار برخوردار بوده و توانسته‌اند عملکرد مناسبی در تشخیص نواحی شکستگی ارائه دهند.

در ادامه، عملکرد سه مدل YOLOv۸، یولو ۹ و یولو ۲۶ بر اساس شاخص‌های استاندارد ارزیابی با یکدیگر مقایسه شد. نتایج نشان داد که هر سه مدل در تشخیص شکستگی‌های مچ دست کودکان به دقت بالایی دست یافته‌اند و اختلاف عملکرد آن‌ها نسبتاً محدود است. این موضوع بیانگر آن است که هر سه معماری از توانایی مناسبی در یادگیری ویژگی‌های مرتبط با شکستگی برخوردار هستند.

بررسی شاخص‌های ارزیابی نشان داد که مدل یولو ۹ در شاخص‌های دقت مثبت، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ بهترین عملکرد را در میان مدل‌های مورد بررسی داشته است، در حالی که مدل یولو ۲۶ در شاخص فراخوانی عملکرد اندکی بهتر از سایر مدل‌ها ارائه داد. با این وجود، اختلاف مقادیر شاخص‌های ارزیابی میان مدل‌ها چشمگیر نبوده و نشان‌دهنده عملکرد مناسب هر سه مدل در مسئله تشخیص شکستگی است.

با توجه به اینکه در این پژوهش، میانگین دقت در بازه آستانه‌های هم‌پوشانی ۰٫۵ تا ۰٫۹۵ به‌عنوان معیار اصلی ارزیابی انتخاب شد، نتایج به‌دست‌آمده نشان می‌دهد که مدل یولو ۹ بهترین عملکرد کلی را در میان مدل‌های مورد بررسی ارائه کرده است. این مدل علاوه بر دستیابی به بالاترین مقدار معیار اصلی ارزیابی، در سایر شاخص‌های کمی نیز عملکردی رقابتی و متوازن از خود نشان داد و به همین دلیل به‌عنوان مناسب‌ترین مدل برای مسئله مورد مطالعه انتخاب شد.

در فصل بعد، یافته‌های این پژوهش در کنار نتایج مطالعات پیشین مورد بحث قرار گرفته، مهم‌ترین دستاوردها، محدودیت‌های پژوهش و پیشنهادهایی برای تحقیقات آینده ارائه خواهد شد.

فصل پنجم جمع‌بندی، نتیجه‌گیری و پیشنهادها

۵-۱ مقدمه

هر پژوهش علمی، علاوه بر تولید نتایج تجربی، نیازمند تفسیر، جمع‌بندی و بررسی میزان تحقق اهداف از پیش تعیین‌شده است. ازاین‌رو، پس از ارائه و تحلیل نتایج حاصل از ارزیابی مدل‌های مورد بررسی، ضروری است یافته‌های پژوهش در چارچوب اهداف و پرسش‌های تحقیق مورد ارزیابی قرار گیرند تا بتوان تصویری جامع از دستاوردهای حاصل و میزان موفقیت روش پیشنهادی ارائه کرد.

هدف این فصل، جمع‌بندی یافته‌های پژوهش، تبیین مهم‌ترین نتایج حاصل از ارزیابی مدل‌ها و ارائه نتیجه‌گیری نهایی بر اساس شواهد تجربی به‌دست‌آمده است. همچنین، در این فصل میزان دستیابی به اهداف پژوهش مورد بررسی قرار گرفته و مهم‌ترین دستاوردها و محدودیت‌های پژوهش بیان می‌شوند. در پایان نیز پیشنهادهایی برای توسعه پژوهش و ادامه مطالعات در این حوزه ارائه خواهد شد.

بدین ترتیب، این فصل ضمن ارائه جمع‌بندی نهایی پژوهش، زمینه لازم برای بهره‌گیری از نتایج حاصل در مطالعات آینده و کاربردهای عملی مرتبط با تشخیص شکستگی‌های مچ دست کودکان مبتنی بر یادگیری عمیق را فراهم می‌سازد.

۵-۲ جمع‌بندی پژوهش

پژوهش حاضر با هدف بررسی توانایی مدل‌های مبتنی بر یادگیری عمیق در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان از تصاویر رادیوگرافی انجام شد. تشخیص سریع و دقیق شکستگی‌های استخوانی، به‌ویژه در بیماران کودک، نقش مهمی در انتخاب روش

درمان، کاهش عوارض ناشی از تشخیص دیرهنگام و بهبود روند درمان دارد. با توجه به پیشرفت‌های اخیر در حوزه بینایی ماشین و یادگیری عمیق، استفاده از مدل‌های آشکارسازی اشیا به‌عنوان ابزارهای کمک‌تشخیصی مورد توجه پژوهشگران قرار گرفته است.

در این پژوهش، به‌منظور دستیابی به ارزیابی قابل اعتماد، از مجموعه‌داده GRAZPEDWRI-DX شامل تصاویر رادیوگرافی مچ دست کودکان استفاده شد. پس از آماده‌سازی داده‌ها، دو مدل یولو ۹ و یولو ۲۶ با استفاده از شرایط آموزشی یکسان آموزش داده شدند. همچنین، به‌منظور تکمیل فرآیند مقایسه، نتایج مدل YOLOv۸ نیز در ارزیابی نهایی مورد استفاده قرار گرفت. یکسان بودن مجموعه‌داده، تنظیمات آموزش و معیارهای ارزیابی برای تمامی مدل‌ها، امکان مقایسه مستقیم عملکرد آن‌ها را فراهم ساخت.

به‌منظور ارزیابی عملکرد مدل‌ها، از معیارهای استاندارد حوزه آشکارسازی اشیا شامل دقت مثبت، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ استفاده شد. علاوه بر این، روند آموزش مدل‌ها، منحنی‌های ارزیابی و ماتریس‌های درهم‌ریختگی نیز مورد بررسی قرار گرفت تا عملکرد مدل‌ها از جنبه‌های مختلف تحلیل شود.

نتایج به‌دست‌آمده نشان داد که هر سه مدل مورد بررسی توانایی مناسبی در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان دارند و اختلاف عملکرد آن‌ها در شاخص‌های ارزیابی نسبتاً محدود است. این موضوع بیانگر آن است که معماری‌های مورد بررسی از قابلیت مناسبی برای استخراج ویژگی‌های مرتبط با شکستگی و آشکارسازی نواحی آسیب‌دیده برخوردار هستند.

در مجموع، پژوهش حاضر ضمن ارائه یک مقایسه منصفانه میان مدل‌های منتخب، تصویری روشن از توانمندی آن‌ها در مسئله تشخیص شکستگی‌های مچ دست کودکان ارائه کرد و زمینه لازم را برای انتخاب مدل مناسب و بررسی قابلیت استفاده از آن در کاربردهای کمک‌تشخیصی فراهم ساخت.

۳-۵ نتیجه‌گیری

نتایج حاصل از این پژوهش نشان داد که استفاده از مدل‌های مبتنی بر یادگیری عمیق، به‌ویژه معماری‌های خانوادهYOLO ، می‌تواند روشی مؤثر برای تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی باشد. عملکرد مناسب مدل‌های مورد بررسی در شاخص‌های استاندارد ارزیابی بیانگر آن است که این معماری‌ها توانایی استخراج ویژگی‌های مؤثر از تصاویر پزشکی و شناسایی نواحی آسیب‌دیده را با دقت قابل قبول دارند.

مقایسه نتایج سه مدل مورد بررسی نشان داد که هر سه مدل عملکرد مناسبی در مسئله تشخیص شکستگی ارائه کرده‌اند و اختلاف آن‌ها در شاخص‌های ارزیابی نسبتاً محدود بوده است. این موضوع بیانگر آن است که تمامی مدل‌های مورد بررسی از قابلیت مناسبی برای استفاده در سامانه‌های کمک‌تشخیصی برخوردار هستند و می‌توانند به‌عنوان ابزارهای پشتیبان تصمیم‌گیری در کنار متخصصان مورد استفاده قرار گیرند.

با وجود عملکرد مناسب هر سه مدل، نتایج ارزیابی نشان داد که یولو ۹ در اغلب شاخص‌های اصلی ارزیابی، از جمله دقت مثبت، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵، عملکرد بهتری نسبت به سایر مدل‌ها ارائه کرده است. از سوی دیگر، مدل یولو ۲۶ در شاخص فراخوانی اندکی عملکرد بهتری از خود نشان داد، اما این برتری با کاهش جزئی در سایر شاخص‌های ارزیابی همراه بود. با توجه به اینکه در این پژوهش، میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵ به‌عنوان معیار اصلی ارزیابی انتخاب شده بود، می‌توان نتیجه گرفت که مدل یولو ۹ مناسب‌ترین عملکرد کلی را در میان مدل‌های مورد بررسی ارائه کرده است.

از سوی دیگر، نزدیک بودن مقادیر شاخص‌های ارزیابی میان مدل‌ها نشان می‌دهد که انتخاب معماری مناسب تنها بر اساس یک معیار امکان‌پذیر نیست و باید مجموعه‌ای از شاخص‌های ارزیابی به‌صورت هم‌زمان مورد توجه قرار گیرند. این موضوع اهمیت استفاده از ارزیابی چندمعیاره در انتخاب مدل‌های یادگیری عمیق برای کاربردهای پزشکی را نشان می‌دهد و بیانگر آن است که تصمیم‌گیری نهایی باید با در نظر گرفتن نیازهای کاربردی، الزامات سامانه و شرایط بهره‌برداری انجام شود.

در مجموع، یافته‌های این پژوهش نشان می‌دهد که مدل‌های مبتنی بر خانواده یولو از ظرفیت مناسبی برای توسعه سامانه‌های هوشمند کمک‌تشخیصی در حوزه تصویربرداری پزشکی برخوردار هستند. همچنین، نتایج حاصل می‌تواند مبنایی برای انجام پژوهش‌های تکمیلی، بهبود معماری مدل‌ها و توسعه سامانه‌های هوشمند تشخیص شکستگی با دقت و قابلیت اطمینان بیشتر باشد.

۵-۴ دستاوردهای پژوهش و کاربردهای عملی

پژوهش حاضر علاوه بر مقایسه عملکرد مدل‌های مبتنی بر یادگیری عمیق در تشخیص شکستگی‌های مچ دست کودکان، دستاوردهایی در حوزه‌های علمی و کاربردی به همراه داشته است. مهم‌ترین دستاورد این پژوهش، ارائه یک چارچوب منسجم برای آموزش، ارزیابی و مقایسه مدل‌های آشکارسازی اشیا در یک مسئله پزشکی، تحت شرایط آموزشی و ارزیابی یکسان است. استفاده از تنظیمات مشترک برای تمامی مدل‌ها موجب شد مقایسه عملکرد آن‌ها بدون تأثیر عوامل جانبی انجام شود و نتایج حاصل، تصویری واقع‌بینانه از توانایی هر معماری ارائه دهد.

از دیگر دستاوردهای این پژوهش، ارزیابی سه مدل از خانواده یولو بر روی مجموعه‌داده GRAZPEDWRI-DX و تحلیل عملکرد آن‌ها با استفاده از معیارهای استاندارد حوزه آشکارسازی اشیا است. استفاده هم‌زمان از شاخص‌های دقت مثبت، فراخوانی، میانگین دقت در آستانه هم‌پوشانی ۰٫۵ و میانگین دقت در بازه آستانه‌های ۰٫۵ تا ۰٫۹۵، امکان ارزیابی جامع عملکرد مدل‌ها را فراهم ساخت و نشان داد که هر یک از آن‌ها دارای نقاط قوت و محدودیت‌های خاص خود هستند.

یافته‌های این پژوهش همچنین نشان داد که مدل‌های مبتنی بر یادگیری عمیق قابلیت مناسبی برای تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان در تصاویر رادیوگرافی دارند و می‌توانند به‌عنوان ابزارهای کمک‌تشخیصی در کنار پزشکان مورد استفاده قرار گیرند.

هرچند چنین سامانه‌هایی جایگزین نظر متخصص نیستند، اما می‌توانند در افزایش سرعت بررسی تصاویر، کاهش احتمال نادیده گرفتن برخی شکستگی‌ها و کمک به فرآیند تصمیم‌گیری بالینی نقش مؤثری ایفا کنند.

علاوه بر این، نتایج به‌دست‌آمده می‌تواند مبنایی برای توسعه پژوهش‌های بعدی در زمینه استفاده از مدل‌های یادگیری عمیق در تصویربرداری پزشکی باشد. چارچوب ارائه‌شده در این پژوهش قابلیت توسعه برای سایر انواع شکستگی‌ها، اندام‌های مختلف و حتی سایر مسائل مرتبط با تحلیل تصاویر پزشکی را نیز دارا است.

به‌طور خلاصه، مهم‌ترین دستاوردهای این پژوهش عبارت‌اند از:

- ارائه چارچوبی منسجم برای آموزش، ارزیابی و مقایسه مدل‌های آشکارسازی اشیا در تشخیص شکستگی‌های مچ دست کودکان .
- مقایسه عملکرد سه مدل از خانواده یولو تحت شرایط آموزشی و ارزیابی یکسان .
- ارزیابی جامع مدل‌ها با استفاده از معیارهای استاندارد و تحلیل نتایج حاصل از آموزش و آزمون .
- شناسایی مناسب‌ترین مدل در میان مدل‌های مورد بررسی بر اساس معیارهای ارزیابی تعریف‌شده در پژوهش .
- ارائه نتایجی که می‌تواند در توسعه سامانه‌های کمک‌تشخیصی مبتنی بر هوش مصنوعی و انجام پژوهش‌های آینده در حوزه تصویربرداری پزشکی مورد استفاده قرار گیرد.

۵-۵ محدودیت‌های پژوهش

پژوهش حاضر، با وجود دستیابی به نتایج مطلوب در تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان، همانند سایر پژوهش‌های حوزه یادگیری عمیق با محدودیت‌هایی همراه بوده است که لازم است در تفسیر نتایج و توسعه پژوهش‌های آینده مورد توجه قرار گیرند.

یکی از مهم‌ترین محدودیت‌های این پژوهش، استفاده از یک مجموعه‌داده برای آموزش و ارزیابی مدل‌ها است. اگرچه این مجموعه‌داده از منابع معتبر و دارای برچسب‌گذاری تخصصی انتخاب شده است، اما استفاده از داده‌های گردآوری‌شده از مراکز درمانی متعدد با شرایط تصویربرداری متفاوت می‌تواند به ارزیابی جامع‌تر میزان تعمیم‌پذیری مدل‌ها کمک کند.

از دیگر محدودیت‌های پژوهش، تمرکز بر یک نوع مسئله تشخیصی است. در این پژوهش، مدل‌ها صرفاً برای تشخیص و مکان‌یابی شکستگی‌های مچ دست کودکان آموزش داده شدند و عملکرد آن‌ها در تشخیص سایر آسیب‌های اسکلتی یا تصاویر مربوط به سایر اندام‌ها مورد بررسی قرار نگرفت. بنابراین، تعمیم نتایج به سایر کاربردهای تصویربرداری پزشکی نیازمند انجام مطالعات تکمیلی است.

همچنین، ارزیابی مدل‌ها بر اساس شاخص‌های استاندارد آشکارسازی اشیا و بر روی داده‌های از پیش برچسب‌گذاری‌شده انجام شد. هرچند این شاخص‌ها معیارهای پذیرفته‌شده‌ای برای سنجش عملکرد مدل‌های یادگیری عمیق هستند، اما ارزیابی عملکرد سامانه در شرایط واقعی محیط‌های درمانی و بررسی تعامل آن با پزشکان متخصص، خارج از دامنه این پژوهش بوده است.

علاوه بر این، در این پژوهش تمرکز اصلی بر مقایسه عملکرد معماری‌های مختلف خانواده یولو بود و سایر خانواده‌های مدل‌های آشکارسازی اشیا یا روش‌های مبتنی بر تقسیم‌بندی تصویر مورد بررسی قرار نگرفتند. ازاین‌رو، نتایج ارائه‌شده صرفاً در چارچوب مدل‌های مورد ارزیابی قابل تفسیر است.

با وجود این محدودیت‌ها، تلاش شد تمامی مدل‌ها تحت شرایط آموزشی و ارزیابی یکسان مورد بررسی قرار گیرند تا نتایج حاصل از مقایسه آن‌ها از حداکثر اعتبار و قابلیت استناد برخوردار باشد.

۶-۵ پیشنهادهایی برای پژوهش‌های آینده

با توجه به نتایج حاصل از این پژوهش و محدودیت‌های مطرح‌شده، می‌توان مسیرهای متعددی را برای توسعه و تکمیل پژوهش‌های آینده پیشنهاد کرد. مهم‌ترین پیشنهادها عبارت‌اند از:

- استفاده از مجموعه‌داده‌های بزرگ‌تر، متنوع‌تر و گردآوری‌شده از مراکز درمانی مختلف، به‌منظور بررسی میزان تعمیم‌پذیری مدل‌ها در شرایط واقعی و کاهش اثر تفاوت‌های ناشی از تجهیزات تصویربرداری و پروتکل‌های مختلف تهیه تصاویر .
- بررسی عملکرد مدل‌های جدیدتر و پیشرفته‌تر حوزه آشکارسازی اشیا و مقایسه آن‌ها با مدل‌های مورد استفاده در این پژوهش، به‌منظور شناسایی معماری‌های مناسب‌تر برای تشخیص شکستگی‌های استخوانی .
- ارزیابی روش‌های مبتنی بر تقسیم‌بندی تصویر و مقایسه آن‌ها با روش‌های مبتنی بر آشکارسازی اشیا، به‌منظور بررسی میزان بهبود دقت در تعیین دقیق ناحیه شکستگی .
- بررسی استفاده از روش‌های یادگیری انتقالی و تنظیم دقیق مدل با استفاده از داده‌های بومی، به‌منظور افزایش دقت مدل‌ها در شرایط بالینی و جمعیت‌های مختلف .
- توسعه مدل‌های هوشمند با قابلیت توضیح‌پذیری، به‌گونه‌ای که علاوه بر ارائه نتیجه تشخیص، دلایل تصمیم‌گیری مدل نیز برای پزشک قابل مشاهده و تفسیر باشد. این موضوع می‌تواند نقش مهمی در افزایش اعتماد متخصصان به سامانه‌های مبتنی بر هوش مصنوعی داشته باشد .
- بررسی امکان ترکیب تصاویر رادیوگرافی با اطلاعات بالینی بیماران، از جمله سن، جنس، سابقه آسیب و علائم بالینی، به‌منظور توسعه سامانه‌های تصمیم‌یار با دقت و قابلیت اطمینان بیشتر .
- ارزیابی عملکرد مدل‌های توسعه‌یافته در محیط‌های درمانی واقعی و بررسی تأثیر آن‌ها بر سرعت تشخیص، کاهش خطاهای انسانی و بهبود فرآیند تصمیم‌گیری بالینی .
- پژوهش در این مسیرها می‌تواند علاوه بر ارتقای دقت سامانه‌های هوشمند تشخیص شکستگی، زمینه را برای توسعه ابزارهای کمک‌تشخیصی قابل اعتمادتر و استفاده گسترده‌تر از فناوری‌های مبتنی بر هوش مصنوعی در حوزه تصویربرداری پزشکی فراهم کند.

منابع

[۱] E. Nagy, M. Janisch, F. Hržić, E. Sorantin, and S. Tschauner, “A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning,” *Scientific Data*, vol. ۹, Art. no. ۲۲۲, ۲۰۲۲, doi: ۱۰,۱۰۳۸/s٤۱٥۹۷-۰۲۲-۰۱۳۲۸-z.

[۲] M. Yaseen, “What is YOLOv۸: An in-depth exploration of the internal features of the next-generation object detector,” *arXiv preprint* arXiv:۲٤۰۸,۱٥۸٥۷, ۲۰۲٤, doi: ۱۰,٤۸٥٥۰/arXiv.۲٤۰۸,۱٥۸٥۷.

[۳] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “یولو ۹: Learning what you want to learn using programmable gradient information,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, ۲۰۲٤, arXiv:۲٤۰۲,۱۳٦۱٦.

[٤] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu, “Ultralytics YOLO۲٦: Unified real-time end-to-end vision models,” *arXiv preprint* arXiv:۲٦۰٦,۰۳۷٤۸, ۲۰۲٦, doi: ۱۰,٤۸٥٥۰/arXiv.۲٦۰٦,۰۳۷٤۸.

[٥] T. Till, S. Tschauner, G. Singer, K. Lichtenegger, and H. Till, “Development and optimization of AI algorithms for wrist fracture detection in children using a freely available dataset,” *Frontiers in Pediatrics*, vol. ۱۱, Art. no. ۱۲۹۱۸۰٤, ۲۰۲۳, doi: ۱۰,۳۳۸۹/fped.۲۰۲۳,۱۲۹۱۸۰٤.

[٦] R.-Y. Ju and W. Cai, “Fracture detection in pediatric wrist trauma X-ray images using YOLOv۸ algorithm,” *Scientific Reports*, vol. ۱۳, Art. no. ۲۰۰۷۷, ۲۰۲۳, doi: ۱۰,۱۰۳۸/s٤۱٥۹۸-۰۲۳-٤۷٤٦۰-۷.

[۷] C.-T. Chien, R.-Y. Ju, K.-Y. Chou, and J.-S. Chiang, “یولو ۹ for fracture detection in pediatric wrist trauma X-ray images,” *Electronics Letters*, vol. ٦۰, no. ۱۱, Art. no. e۱۳۲٤۸, ۲۰۲٤, doi: ۱۰,۱۰٤۹/ell۲,۱۳۲٤۸.

[۸] J. Jung, J. Dai, B. Liu, and Q. Wu, “Artificial intelligence in fracture detection with different image modalities and data types: A systematic review and meta-analysis,” *PLOS Digital Health*, vol. ۳, no. ۱, Art. no. e۰۰۰۰٤۳۸, ۲۰۲٤, doi: ۱۰,۱۳۷۱/journal.pdig.۰۰۰۰٤۳۸.

[۹] A. Nowroozi, M. A. Salehi, P. Shobeiri, S. Agahi, S. Momtazmanesh, P. Kaviani, and M. K. Kalra, “Artificial intelligence diagnostic accuracy in fracture detection from plain radiographs and comparing it with clinicians: A systematic review and meta-analysis,” *Clinical Radiology*, vol. ۷۹, no. ۸, pp. ٥۷۹–٥۸۸, ۲۰۲٤, doi: ۱۰,۱۰۱٦/j.crad.۲۰۲٤,۰٤,۰۰۹.

[۱۰] K. Suen, R. Zhang, and N. Kutaiba, “Accuracy of wrist fracture detection on radiographs by artificial intelligence compared to human clinicians: A systematic review and meta-analysis,” *European Journal of Radiology*, vol. ۱۷۸, Art. no. ۱۱۱٥۹۳, ۲۰۲٤, doi: ۱۰,۱۰۱٦/j.ejrad.۲۰۲٤,۱۱۱٥۹۳.

[١١] M. Tariq and K. Choi, “YOLO١١-driven deep learning approach for enhanced detection and visualization of wrist fractures in X-ray images,” *Mathematics*, vol. ١٣, no. ٩, Art. no. ١٤١٩, ٢٠٢٥, doi: ١٠,٣٣٩٠/math١٣٠٩١٤١٩.

[١٢] M. Scherkl, N. Stranger, A. Ciornei-Hoffman, G. Singer, T. Till, H. Till, F. Hržić, and S. Tschauner, “Automated AI fracture detection in initial presentation pediatric wrist X-rays: Effects and benefits of adding follow-up examinations,” *La Radiologia Medica*, vol. ١٣١, no. ٣, pp. ٤٥٨–٤٦٩, ٢٠٢٦, doi: ١٠,١٠٠٧/s١١٥٤٧-٠٢٥-٠٢١٥٣-١.

[١٣] J. R. Zech, G. Carotenuto, Z. Igbinoba, C. V. Tran, E. Insley, A. Baccarella, and T. T. Wong, “Detecting pediatric wrist fractures using deep-learning-based object detection,” *Pediatric Radiology*, vol. ٥٣, no. ٦, pp. ١١٢٥–١١٣٤, ٢٠٢٣, doi: ١٠,١٠٠٧/s٠٠٢٤٧-٠٢٣-٠٥٥٨٨-٨.

[١٤] A. Ahmed, A. S. Imran, A. Manaf, Z. Kastrati, and S. M. Daudpota, “Enhancing wrist fracture detection with YOLO,” *arXiv preprint* arXiv:٢٤٠٧,١٢٥٩٧, ٢٠٢٤.