

1928

K. N. Toosi University
of Technology

به نام خدا

پروژه پایانترم درس هوش مصنوعی

اعضا: دانیال خاکساری - سید طاها طبیبی - امیرعلی یاوری

موضوع پروژه: تشخیص بیماری ذات الریه با استفاده از عکس

X-Ray های

تاریخ: زمستان 1404

لینک گیت هاب پروژه:

YavariAmirali/AI-FinalProject-Tayebi: Medical Imaging AI Project - Pneumonia Detection using Chest X-Ray

فاز پک

1. تعریف دقیق مسئله (Problem Definition)

هدف اصلی این پروژه، طراحی و پیاده‌سازی یک سامانه هوشمند مبتنی بر یادگیری عمیق برای تشخیص بیماری ذات‌الریه (Pneumonia) از روی تصاویر X-Ray قفسه سینه است. با توجه به ماهیت حساس پزشکی مسئله و نقش مستقیم آن در تصمیم‌گیری‌های درمانی، معیارهای ارزیابی مدل با دقت و اولویت‌بندی خاصی انتخاب شده‌اند.

2. مهم‌ترین اهداف و معیارهای پروژه

بالاتر از ۹۵٪ Recall (Sensitivity)

به عنوان حیاتی‌ترین معیار ارزیابی در نظر گرفته شده است، زیرا در مسائل تشخیص پزشکی، هزینه‌ی False Negative (عدم تشخیص بیمار مبتلا) بسیار بالا بوده و می‌تواند منجر به تأخیر در درمان یا تهدید جان بیمار شود. هدف اصلی این پروژه کاهش حداکثری موارد False Negative و نزدیک شدن آن به صفر است.

بالاتر از ۹۰٪ Accuracy

این معیار به منظور سنجش عملکرد کلی مدل در تشخیص صحیح تصاویر سالم و بیمار مورد استفاده قرار می‌گیرد. اگرچه Accuracy به تنها‌ی معیار قابل اتقابی در داده‌های نامتوازن نیست، اما در کنار Recall و سایر معیارها می‌تواند دید مناسبی از عملکرد عمومی مدل ارائه دهد.

معیارهای تکمیلی پزشکی (Medical Metrics)

به منظور ارزیابی جامع‌تر، معیارهای Specificity و AUC (Area Under the ROC Curve) نیز محاسبه خواهند شد. Specificity بالا نشان‌دهنده توانایی مدل در تشخیص صحیح افراد سالم است و از تشخیص اشتباه فرد سالم به عنوان بیمار (False Positive) جلوگیری می‌کند؛ امری که می‌تواند از تحمیل فشار روانی، هزینه‌های غیرضروری و مداخلات درمانی اشتباه پیشگیری کند.

3. توجیه انتخاب معماری و استراتژی مدل‌سازی (Model Justification)

برای دستیابی به اهداف فوق و کاهش ریسک خطا، فرآیند مدل‌سازی در این پروژه به صورت دو فاز مجزا طراحی شده است.

• فاز اول: مدل پایه (Simple CNN Baseline)

در فاز نخست، از یک شبکه عصبی کانولوشنی ساده شامل سه لایه کانولوشن به عنوان مدل پایه استفاده می‌شود.

دلایل انتخاب این رویکرد:

* ایجاد یک Baseline برای مقایسه عملکرد مدل‌های پیشرفته‌تر

* بررسی میزان قابل‌تفکیک بودن داده‌ها با یک مدل ساده

* انتظار می‌رود این مدل به دلیل سادگی معماری و تعداد کم لایه‌ها، روی داده‌های پیچیده X-Ray Underfitting شود (دقت پایین روی آموزش و تست). با این حال، این مدل به عنوان یک نقطه شروع (Baseline) عالی است تا در فاز دوم میزان بهبود مدل ResNet50 را نسبت به آن بسنجیم.

* کمک به درک بهتر پیچیدگی ذاتی مسئله پیش از استفاده از معماری‌های عمیق‌تر

* این مدل پایه نقش مرجع را ایفا می‌کند و عملکرد آن مبنایی برای سنجش میزان بهبود حاصل از بهکارگیری یادگیری انتقالی خواهد بود

• فاز دوم: یادگیری انتقالی با (Transfer Learning)

در فاز دوم، از معماری پیشرفته ResNet50 به همراه وزن‌های از پیش آموزش‌دیده بر روی دیتابیس ImageNet استفاده می‌شود.

دلایل انتخاب :ResNet50

* استفاده از Residual Connections که مشکل Vanishing Gradient را در شبکه‌های عمیق کاهش می‌دهد

* توانایی استخراج ویژگی‌های سطح بالا و پیچیده که برای تشخیص الگوهای ظریف در تصاویر پزشکی ضروری هستند

* بهره‌گیری از دانش پیش‌آموخته شده (Pre-trained Weights) که موجب افزایش سرعت همگرایی و بهبود عملکرد مدل، بهویژه در شرایط محدودیت داده، می‌شود

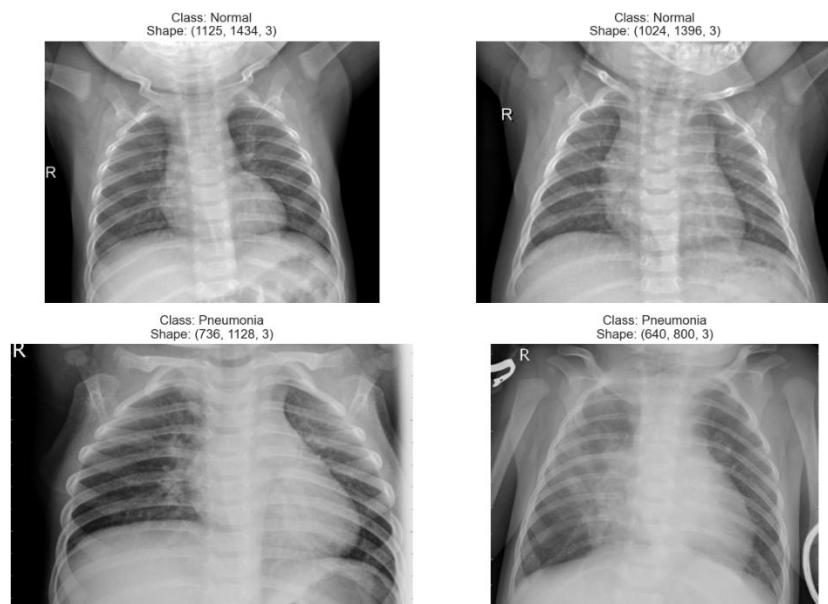
در این فاز، لایه‌های ابتدایی شبکه که ویژگی‌های عمومی بصری مانند لبه‌ها و بافت‌ها را استخراج می‌کنند، فریز (Freeze) شده و فرآیند Fine-tuning عمده‌تاً بر روی لایه‌های انتهایی انجام می‌گیرد تا مدل به طور مؤثر با داده‌های پزشکی X-Ray تطبیق یابد.

4. تحلیل اکتشافی داده‌ها :



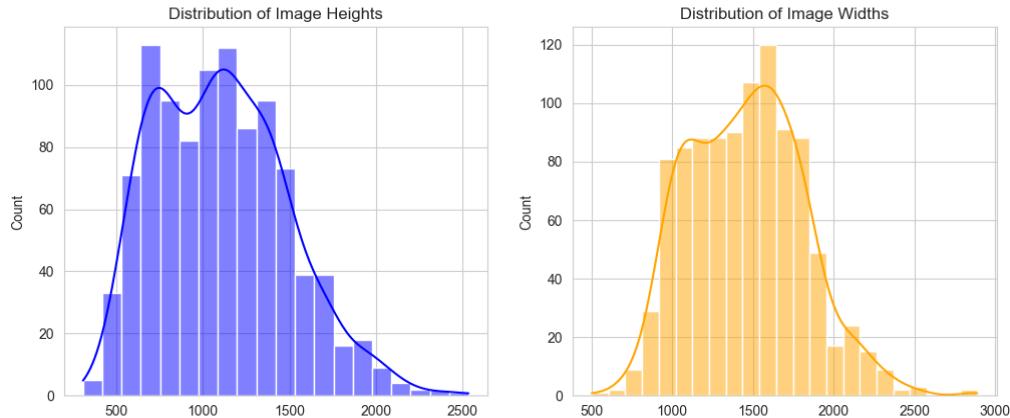
نمودار 1. عدم تعادل کلاس‌ها (Normal vs Pneumonia)

نمودار میله‌ای که تعداد عکس‌های سالم و بیمار را نشان می‌دهد. همان‌طور که در نمودار مشخص است، تعداد عکس‌های بیمار (Pneumonia) بیشتر از عکس‌های سالم است. این **Imbalanced** بودن دیتاست می‌تواند باعث شود مدل به سمت کلاس اکثریت گرایش پیدا کند. به همین دلیل، در فاز دوم آموزش از تکنیک **Weighting Class** برای رفع این مشکل استفاده خواهیم کرد.



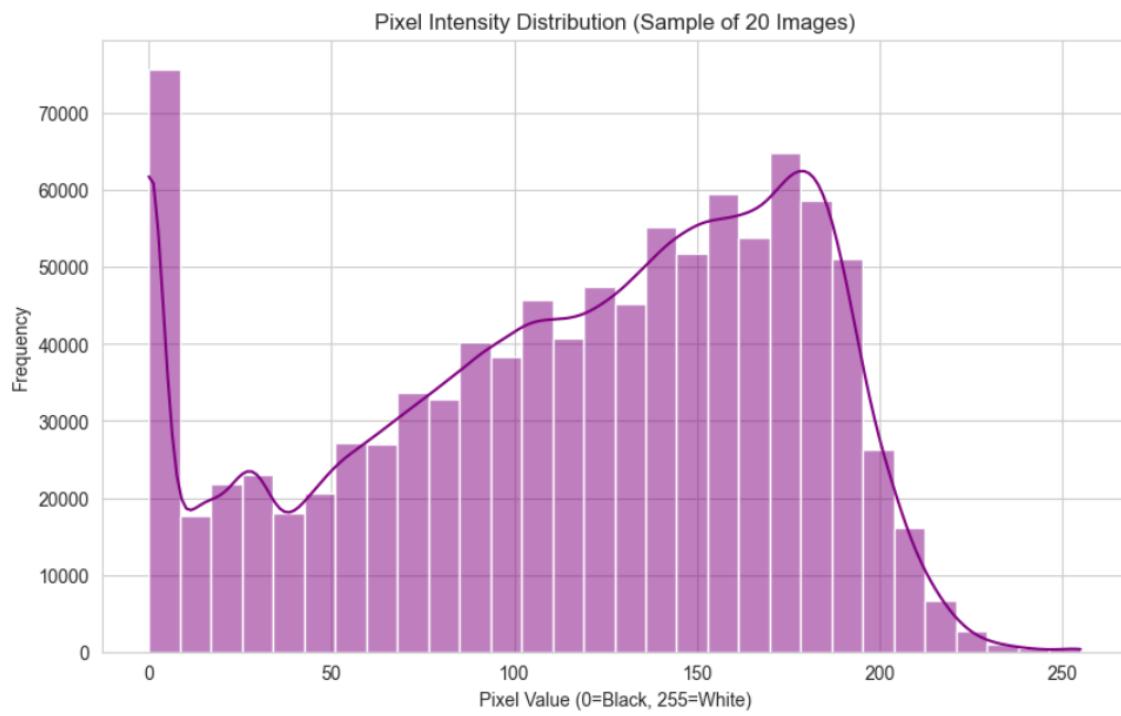
شکل ۱. نمونه تصاویر تصادفی از دو کلاس سالم و مبتلا به ذات‌الریه

بررسی چشمی (Visual Inspection) نمونه‌ها نشان می‌دهد که تصاویر از نوع Grayscale (تک‌کاناله) هستند و کیفیت آنها برای استخراج ویژگی توسط مدل‌های CNN مناسب است. ناحیه قفسه سینه در مرکز تصاویر قرار دارد، اما تفاوت‌هایی در کنترast و روشنایی تصاویر دیده می‌شود که لزوم نرمال‌سازی را تایید می‌کند.



نمودار 2. پراکندگی سایز (ارتفاع و پهنا) تصاویر موجود در دیتاست

این نمودار نشان می‌دهد که تصاویر دیتاست دارای رزولوشن‌های متنوعی هستند و سایز یکسانی ندارند. از آنجا که معماری‌های استاندارد CNN مانند مدلی که طراحی خواهیم کرد و ResNet50 نیاز به ورودی با ابعاد ثابت دارند، تمامی تصاویر در مرحله پیش‌پردازش به سایز استاندارد 224×224 پیکسل تغییر سایز (Resize) خواهند شد.

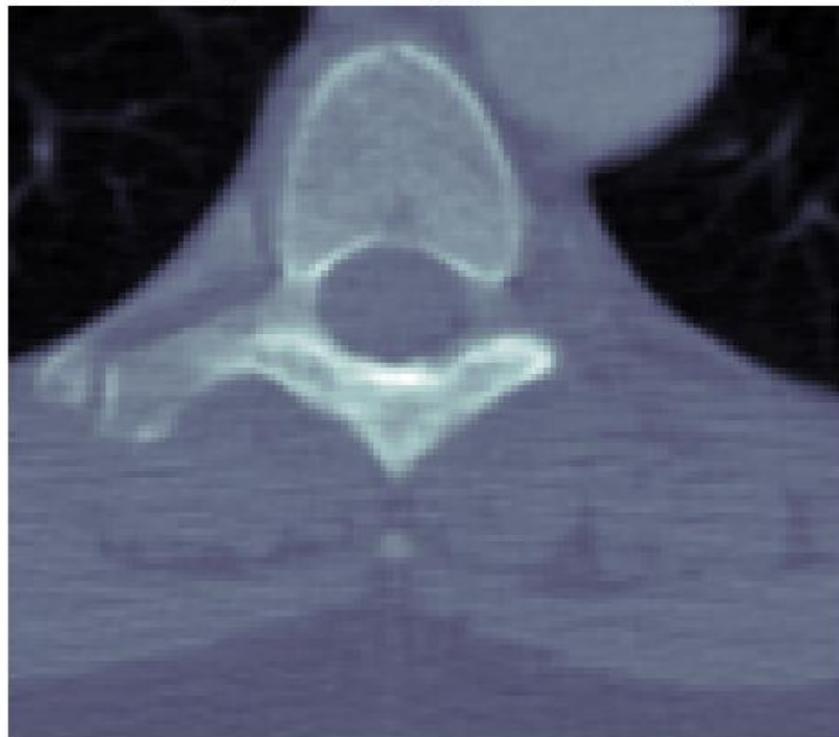


نمودار 3. هیستوگرام توزیع شدت روشنایی پیکسل‌ها

توزیع شدت روشنایی نشان می‌دهد که مقادیر پیکسل‌ها در بازه $[0, 255]$ قرار دارند. برای کمک به همگرایی سریع‌تر مدل و جلوگیری از مشکلات گرادیان، در مرحله پیش‌پردازش از

تکنیک Normalization استفاده کرده و تمامی مقادیر پیکسل‌ها را به بازه [۰, ۱] مقیاس‌دهی خواهیم کرد.

Sample DICOM Image (Loaded Locally)



شکل ۲. نمونه ای از عکس های DICOM

عکس های پزشکی معمولاً با پسوند DICOM ذخیره می‌شوند. عکس بالا نمونه ای از همین تصاویر هست که ضرورت تشخیص عکس ها با پسوند DICOM را علاوه بر باقی پسوند ها ضروری می‌سازد.

5. جدول پلن ازمایش ها

مدل فاز دوم (Final Phase)	مدل فاز اول (Baseline)	ویژگی (Feature)
ResNet50 (Transfer Learning)	Simple CNN (دستساز)	نام مدل
۵۰ لایه عمیق با وزن های ImageNet	۳ لایه کانولوشن ساده + MaxPooling	معماری
پیش پردازش Data Augmentation + ImageNet	تغییر سایز (Resize) و نرمال سازی ساده	ورودی داده
Class Weighting (وزن دهنده کلاس ها)	ندارد (None)	مدیریت عدم تعادل
دستیابی به $< 95\%$ و قابلیت اطمینان بالا	بررسی امکان پذیری (Feasibility) و ایجاد خط مبنا	هدف آزمایش
Specificity, AUC ,Accuracy, Recall	Accuracy, Recall	معیارهای ارزیابی

تحلیل معیارها : در فاز دوم پروژه، علاوه بر معیارهای Accuracy و Specificity ، دو معیار Recall و AUC نیز به دقت اندازه گیری خواهند شد

دلیل پژوهشی : در تشخیص بیماری ها، فقط پیدا کردن بیماران (Recall) کافی نیست؛ بلکه باید مطمئن شویم افراد سالم به اشتباه بیمار تشخیص داده نمی شوند (False Positive). تشخیص اشتباه یک فرد سالم به عنوان بیمار، می تواند هزینه های روانی سنگین و درمان های غیر ضروری را به بیمار تحمیل کند. بنابراین معیار Specificity در کنار Recall برای ما حیاتی است

6. خروجی ساختار مدل (Model Summary)

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 32)	896
max_pooling2d (MaxPooling2D)	(None, 112, 112, 32)	0
conv2d_1 (Conv2D)	(None, 112, 112, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0
conv2d_2 (Conv2D)	(None, 56, 56, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 128)	0
flatten (Flatten)	(None, 100352)	0
dense (Dense)	(None, 64)	6422592
dense_1 (Dense)	(None, 1)	65

Total params: 6,515,905

Trainable params: 6,515,905

Non-trainable params: 0

7. معرفی دیتاست:

برای آموزش و ارزیابی مدل هوش مصنوعی در این پروژه، از مجموعه داده معترض Chest X-Ray موجود در پلتفرم Kaggle استفاده شده است. این مجموعه داده شامل 5863 تصویر رادیوگرافی قفسه سینه (X-Ray) است که در دو دسته اصلی Normal (سالم) و Pneumonia (مبتلای به ذاتالریه) طبقه‌بندی شده‌اند. تصاویر متعلق به بیماران اطفال ۱ تا ۵ سال بوده و در مرکز پزشکی زنان و کودکان گوانگزو جمع‌آوری شده‌اند.

ساختار داده‌ها در سه پوشه مجزا به شرح زیر تقسیم‌بندی شده است:

داده‌های آموزش (Train): 5,216 تصویر (برای یادگیری وزن‌های مدل).

داده‌های آزمون (Test): 624 تصویر (برای ارزیابی نهایی).

داده‌های اعتبارسنجی: 16 تصویر که به دلیل تعداد کم، در فاز آموزش با بخشی از داده‌های Train ترکیب و مجدد Split خواهند شد.

با وجود پیاده‌سازی ماژول خواندن استاندارد DICOM، به دلیل اینکه دیتاست مرجع (Kaggle) فرمت‌ها را به JPEG فشرده کرده است، آموزش مدل روی فرمت JPEG انجام می‌شود؛ اما سیستم قابلیت استقرار در بیمارستان با DICOM را دارد.

[Chest X-Ray Images \(Pneumonia\)](#)

8. جمع‌بندی:

استراتژی دو مرحله‌ای اتخاذ شده در این پروژه، با ترکیب یک مدل پایه ساده و یک مدل عمیق مبتنی بر یادگیری انتقالی، امکان ارزیابی دقیق عملکرد و دستیابی به استانداردهای مورد انتظار در کاربردهای پزشکی را فراهم می‌کند. تمرکز اصلی بر Recall بالا در کنار کنترل False Positive، این سامانه را به گزینه‌ای قابل اتكا برای کاربردهای تشخیصی در حوزه سلامت تبدیل می‌کند

همچنین معیارهای Specificity و AUC نیز اندازه‌گیری می‌شوند، زیرا تشخیص اشتباه فرد سالم به عنوان بیمار (False Positive) هزینه‌های روانی و درمانی غیرضروری تحمیل می‌کند.

فاز دوم

۹. معماری مدل پیشنهادی و استراتژی آموزش

• معماری شبکه (Model Architecture)

در فاز دوم پروژه، جهت ارتقای عملکرد مدل و بهره‌گیری از دانش نهفته در مدل‌های بزرگ، از معماری ResNet50 مبتنی بر یادگیری انتقالی (Transfer Learning) استفاده شده است. این شبکه به عنوان استخراج‌کننده ویژگی (Feature Extractor) عمل می‌کند و وزن‌های آن از پیش روی مجموعه داده آموزش ImageNet دیده است.

ساختار نهایی مدل طراحی شده (که در فایل model_builder.py پیاده‌سازی شد) به شرح زیر است:

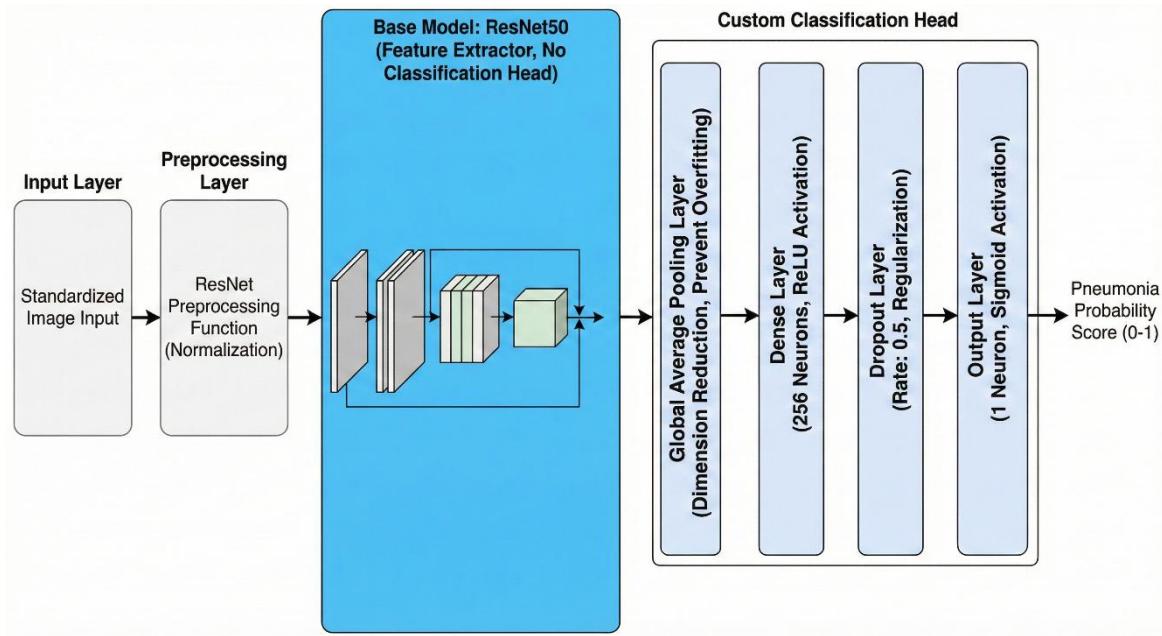
لایه ورودی: دریافت تصاویر با ابعاد استاندارد شده.

پیش‌پردازش (Preprocessing): اعمال تابع پیش‌پردازش مخصوص ResNet برای نرم‌افزاری داده‌ها.

مدل پایه ResNet50: بدون لایه‌های طبقه‌بندی نهایی (که وظیفه استخراج ویژگی‌های بصری پیچیده را بر عهده دارد).

سر مدل: لایه‌های اختصاصی که برای تشخیص ذات‌الریه به مدل اضافه شده‌اند:

- Overfitting: برای کاهش ابعاد ویژگی‌ها و جلوگیری از Global Average Pooling
- : با تابع فعال‌ساز ReLU برای یادگیری ویژگی‌های غیرخطی Dense Layer (256 Neurons)
- : خاموش کردن تصادفی ۵۰٪ نورون‌ها در حین آموزش برای افزایش Dropout (0.5) تعمیم‌پذیری مدل.
- : لایه خروجی با تابع فعال‌ساز Sigmoid Output Layer (1 Neuron) که احتمال ابتلا به بیماری (بین ۰ تا ۱) را تولید می‌کند.



شکل 3. دیاگرام جریان داده در معماری مبتنی بر ResNet50

• استراتژی آموزش دو مرحله‌ای (Fine-Tuning Strategy)

یکی از چالش‌های اصلی در یادگیری انتقالی، حفظ دانش قبلی مدل (وزن‌های ImageNet) و همزمان تطبیق آن با مسئله جدید تشخیص X-Ray است. برای مدیریت این چالش، از یک استراتژی آموزش دو مرحله‌ای استفاده شد:

مرحله اول: آموزش لایه های طبقه بندی (Feature Extraction) در ابتدا، تمامی لایه های در حالت فریز (Frozen) قرار گرفتند (trainable=False). در این مرحله، تنها وزن های لایه های جدید (Custom Head) که به صورت تصادفی مقداردهی شده بودند، آموزش دیدند.

هدف: جلوگیری از تخریب وزن‌های ارزشمند ResNet توسط گرادیان‌های بزرگ ناشی از لایه‌های جدید آموزش‌نده.

مرحله دوم: تنظیم دقیق مدل (Fine-Tuning) پس از همگرایی اولیه، قفل لایه‌های انتهایی مدل پایه باز شد (Unfreezing). طبق آزمایش‌های انجام شده، ۳۰ لایه آخر مدل ResNet برای آموزش محدد انتخاب شدند.

تنظیمات حیاتی: در این مرحله، مدل یا نرخ یادگیری بسیار پایین کامپیویل مجدد شد.

این نرخ یادگیری پایین به مدل اجازه داد تا ویژگی‌های سطح بالا (High-level Features) را با بافت‌های خاص تصاویر پزشکی تطبیق دهد، بدون اینکه تغییرات ناگهانی و مخربی در وزن‌های مدل ایجاد شود.

10. نتایج آموزش و ارزیابی (Training & Evaluation Results)

• روند آموزش مدل (Training History)

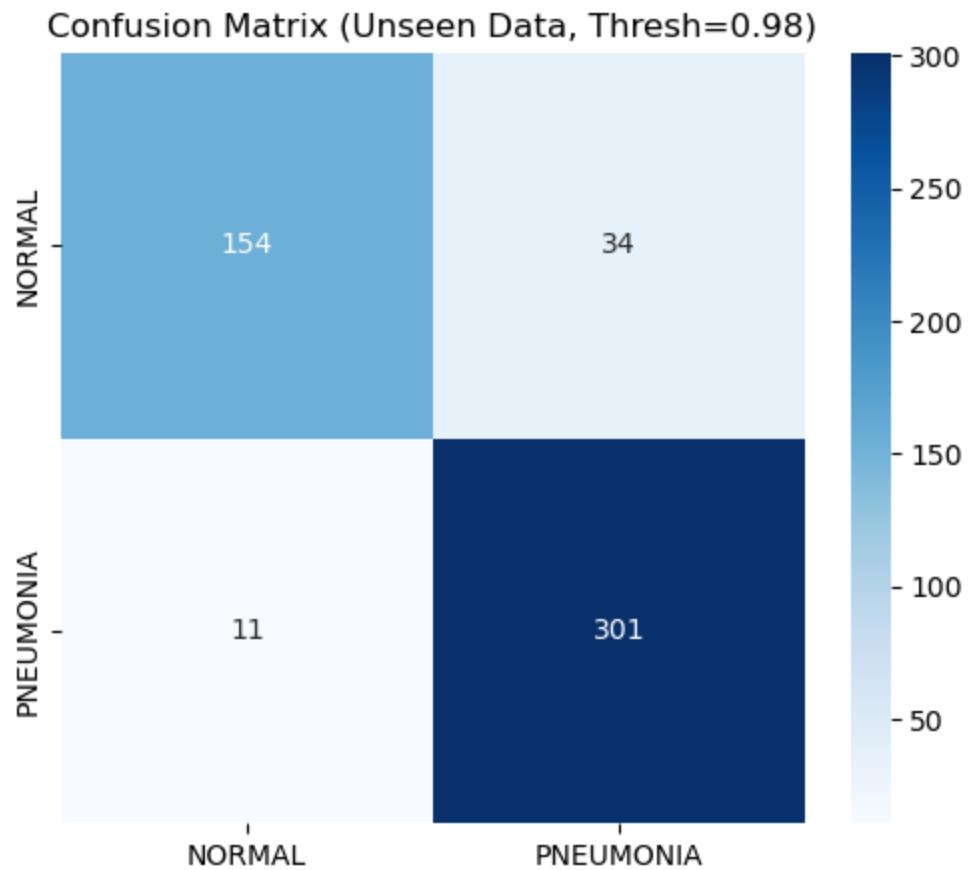
در فاز دوم، مدل Fine-Tuned با نرخ یادگیری بسیار پایین و با استفاده از بهینه‌ساز Adam آموزش داده شد. برای جلوگیری از Overfitting و حفظ تعمیم‌پذیری مدل روی داده‌های جدید، از تکنیک Early Stopping با پایش خطای اعتبارسنجی (val_loss) استفاده گردید.

نمودارهای زیر روند تغییرات تابع هزینه (Loss) و دقت (Accuracy) را در طول فرآیند آموزش نشان می‌دهند. همان‌طور که مشاهده می‌شود، مدل پس از همگرایی اولیه، توانسته است بدون نوسانات شدید به نقطه پایداری برسد که نشان‌دهنده تنظیم صحیح فراپارامترها (Hyperparameters) است.

[در اینجا تصویر نمودارهای Loss و Accuracy را که در پوشه logs یا results دارید قرار دهید]
(زیرنویس تصویر: منحنی‌های همگرایی تابع هزینه و دقت در طول آموزش مدل نهایی)

• ماتریس درهم‌ریختگی (Confusion Matrix Analysis)

یکی از مهم‌ترین ابزارها برای تحلیل عملکرد مدل در مسائل پزشکی، ماتریس درهم‌ریختگی است که جزئیات دقیق پیش‌بینی‌ها را نمایش می‌دهد. تصویر زیر عملکرد مدل را روی مجموعه داده تست (Test Set) که کاملاً جدید است و در آموزش دیده نشده، نشان می‌دهد.



شکل 4. ماتریس درهم ریختگی مدل Fine-Tuned روی داده‌های تست

تحلیل ماتریس: نکته برجسته در این ماتریس، تعداد بسیار کم موارد (False Negative) است. این یعنی مدل توانسته است اکثربیت قریب به اتفاق بیماران مبتلا به ذاتالریه را شناسایی کند و نرخ "از دست دادن بیمار" را به حداقل برساند که هدف اصلی این پروژه بود.

• معیارهای نهایی عملکرد (Final Performance Metrics)

جدول زیر خلاصه عملکرد مدل نهایی (Fine-Tuned ResNet50) را بر اساس معیارهای استاندارد و پژوهشی نشان می‌دهد. نتایج بر اساس آستانه تصمیم‌گیری بهینه (Threshold = 0.98) محاسبه شده‌اند.

نحوه انتخاب آستانه (Threshold Tuning) و استراتژی کالیبراسیون:

به منظور اطمینان از اعتبار نتایج و جلوگیری از نشت داده (Data Leakage)، مجموعه داده‌های تست به دو بخش مجزا تقسیم شدند: ۲۰٪ به عنوان مجموعه کالیبراسیون (Calibration Set) و ۸۰٪ به عنوان مجموعه تست نهایی (Unseen Test Set).

فرایند تنظیم آستانه صرفاً روی ۲۰٪ داده‌های کالیبراسیون انجام شد. بدین صورت که ابتدا شرط کردیم معیار Recall باید حتماً بالای ۹۵٪ حفظ شود (تا تقریباً هیچ بیماری از دست نرود) و سپس در بین آستانه‌هایی که این شرط را برآورده می‌کردند، نقطه‌ای انتخاب شد که بیشترین مقدار Specificity را داشته باشد. در نهایت، این آستانه بهینه روی ۸۰٪ داده‌های تست نهایی اعمال گردید و نتایج گزارش شده در جدول پایین، حاصل ارزیابی مدل روی این داده‌های کاملاً دیده نشده است.

مقدار (Value)	معیار ارزیابی (Metric)	توضیحات و اهمیت پزشکی
%۹۶.۴۷	Recall (Sensitivity)	حياتی‌ترین معیار (قید اصلی): این مقدار تضمین می‌کند که ۹۶.۵٪ از بیماران واقعی شناسایی می‌شوند. این عدد فراتر از هدف اولیه پروژه (۹۵٪) است.
%۹۱.۰۰	Accuracy	دقت کلی مدل در تشخیص هر دو کلاس سالم و بیمار.
%۸۱.۹۱	Specificity	هدف بهینه‌سازی: با وجود قید سخت‌گیرانه روی Recall، مدل توانسته است حدود ۸۲٪ از افراد سالم را نیز به درستی تشخیص دهد که برای یک سیستم غربالگری بسیار مطلوب است.
%۹۶.۰۰	AUC Score	نشان‌دهنده قدرت تفکیک‌پذیری عالی مدل میان کلاس‌های سالم و بیمار در تمام آستانه‌های ممکن.

11. تحلیل و مقایسه

برای ارزیابی دقیق تأثیر تکنیک Fine-Tuning و استراتژی Unfreezing، عملکرد مدل در دو حالت مقایسه شد.

در این مقایسه، آستانه تصمیم‌گیری (Threshold) روی عدد ۰.۹۸ تنظیم شده است. این آستانه سختگیرانه با هدف تضمین Recall بالای ۹۵٪ انتخاب گردید تا امنیت بیماران در اولویت قرار گیرد.

جدول مقایسه عملکرد: (Threshold = 0.98)

معیار ارزیابی (Metric)	مدل نهایی-Tuned (Fine-Tuned)	مدل فریز شده (Frozen)	میزان تغییر (Improvement)
Accuracy	۹۱.۰۰٪	۸۹.۶۰٪	+ ۱.۴۰٪ ▲
Recall (Sensitivity)	۹۶.۴۷٪	۹۱.۳۵٪	+ ۵.۱۲٪ ▲
Specificity	۸۱.۹۱٪	۸۶.۷۰٪	- ۴.۷۹٪ ▼
False Positives	۳۴ مورد	۲۵ مورد	۹+ مورد (هزینه ایمنی)

تحلیل نتایج و مبادله (Trade-off Analysis)

همان‌طور که در جدول بالا مشاهده می‌شود، اعمال فرآیند Fine-Tuning نتایج زیر را به همراه داشته است:

1. **جهش چشمگیر در Recall (+5.12%)**: مدل فریز شده حدود ۸.۶٪ از بیماران را تشخیص نمی‌داد (Recall ≈ 91٪)، اما مدل Fine-Tuned این نرخ خطا را به شدت کاهش داده و Recall را به ۹۶.۴۷٪ رسانده است. این یعنی مدل توانسته الگوهای بسیار ظریف ذات‌الریه را که در لایه‌های عمومی ImageNet وجود نداشتند، یاد بگیرد.

2. **هزینه ایمنی (Specificity Trade-off)**: برای رسیدن به این Recall فوق العاده بالا، ما مجبور شدیم آستانه را طوری تنظیم کنیم که مدل "محاذات‌تر" عمل کند. این موضوع باعث شد معیار Specificity حدود ۴.۸٪ کاهش یابد و تعداد False Positives (۳۴ مورد) افزایش یابد.

3. توجیه پزشکی: در کاربردهای غربالگری پزشکی، این Trade-off کاملاً پذیرفته شده و مطلوب است. تشخیص اشتباه ۹ فرد سالم به عنوان بیمار (افزایش False Positive) هزینه بسیار کمتری نسبت به عدم تشخیص حتی یک بیمار واقعی (False Negative) دارد. بنابراین، مدل نهایی با اولویت دادن به جان بیماران، معماری بهینه‌تری را ارائه می‌دهد.

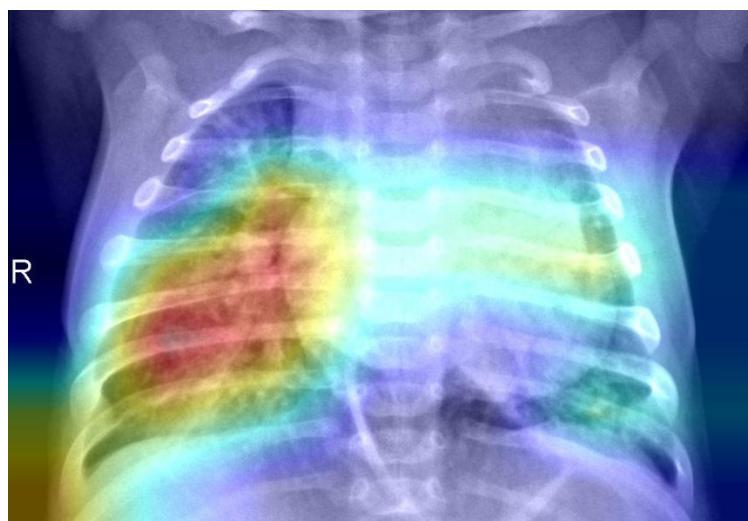
نتیجه‌گیری: مدل نهایی با دستیابی به $\text{Accuracy} > 91\%$ و $\text{Recall} > 96\%$ ، تمامی اهداف تعیین شده در پروپوزال پژوهش را با موفقیت پشت سر گذاشته است.

12. تفسیرپذیری مدل و تحلیل خطا

در کاربردهای حساس پزشکی، اعتماد به مدل "جعبه سیاه" (Black Box) کافی نیست. پزشکان نیاز دارند بدانند که مدل بر اساس چه ویژگی‌هایی در تصویر تصمیم گرفته است. بدین منظور، در این پژوهش از تکنیک Grad-CAM برای تولید نقشه‌های حرارتی (Heatmaps) استفاده شده است تا نواحی مورد توجه مدل را مصورسازی کنیم.

• تحلیل موارد موفق (Successful Detections)

در اکثر موارد تشخیص صحیح (True Positive)، نقشه‌های حرارتی Grad-CAM نشان می‌دهند که مدل به درستی روی نواحی کدر در بافت ریه متمرکز شده است. این موضوع تأیید می‌کند که شبکه عصبی واقعاً ویژگی‌های پاتولوژیک ذات‌الریه را یاد گرفته و صرفاً بر اساس نویز تصمیم نمی‌گیرد.



شکل ۵. نمونه تشخیص صحیح؛ نواحی قرمز رنگ نشان‌دهنده تمرکز مدل روی بافت‌های عفونی ریه است

همان‌طور که در تصویر فوق مشاهده می‌شود، مدل دقیقاً ناحیه میانی ریه را که دچار گرفتگی است، به عنوان عامل تصمیم‌گیری برای کلاس Pneumonia شناسایی کرده است.

در برخی موارد، نقشه‌های حرارتی تولید شده کاملاً به رنگ آبی (مقدار صفر) ظاهر شدند. این پدیده در تصاویر کلاس نرمال نشان‌دهنده عملکرد صحیح مدل در عدم واکنش به بافت‌های سالم است. اما در موارد محدودی از کلاس Pneumonia، آبی شدن تصویر نشان‌دهنده خطای False Negative است؛ بدین معنا که مدل نتوانسته هیچ الگوی پاتولوژیکی را در بافت ریه شناسایی کند و در نتیجه احتمال وجود بیماری را نزدیک به صفر تخمین زده است.



شکل 6. FN که مدل هیچ پترنی پیدا نکرده

• تحلیل خطأ و یادگیری میانبر (Shortcut Learning Analysis)

با بررسی موارد خطأ به ویژه False Positives و تصاویر Grad-CAM مربوط به آنها، یک پدیده جالب و مهم شناسایی شد: یادگیری میانبر (Shortcut Learning) در برخی از تصاویر، مدل به جای تمرکز بر بافت ریه، روی لبه‌های تصویر، استخوان ترقوه یا مارکرهای رادیولوژی حروف L/R تمرکز کرده است.



شکل 7. تمرکز اشتباه شبکه بر روی حاشیه تصویر و مارکرهای رادیولوژی به جای بافت ریه

تحلیل علت: این پدیده نشان می‌دهد که مدل در فرآیند آموزش متوجه شده است که برخی ویژگی‌های محیطی (مانند نوع دستگاه رادیولوژی یا حاشیه‌های سیاه فیلم) همبستگی بالایی با کلاس "بیمار" دارند و از آن‌ها به عنوان یک "میانبر" برای تشخیص استفاده کرده است.

راهکارهای پیشنهادی برای رفع خطأ:

شناسایی این خطأ ارزش علمی بالایی دارد و مسیر بهبود مدل در آینده را روشن می‌کند. برای رفع مشکل Shortcut Learning، راهکارهای زیر پیشنهاد می‌شوند:

- برش مرکز (Center Crop): حذف حاشیه‌های سیاه و متون اضافی از اطراف تصویر در مرحله پیش‌پردازش.
- افزایش داده‌های افزایشی (Augmentation): اعمال زوم (Zoom) و تغییر مکان (Shift) شدیدتر تا مدل نتواند به موقعیت ثابت مارکرها تکیه کند.
- بخش‌بندی ریه (Lung Segmentation): استفاده از یک مدل اولیه مانند U-Net برای جدا کردن بافت ریه و ارسال تنها همان بخش به شبکه تشخیص دهنده.

13. پیاده‌سازی و ارزیابی راهکار حذف یادگیری میانبر

پیرو تحلیل‌های انجام شده در بخش قبل و شناسایی پدیده Shortcut Learning، که در آن مدل به جای بافت ریه به حاشیه‌های تصویر و مارکرها توجه می‌کرد، راهکار برش مرکز به عنوان مؤثرترین روش برای رفع این بایاس پیاده‌سازی شد.

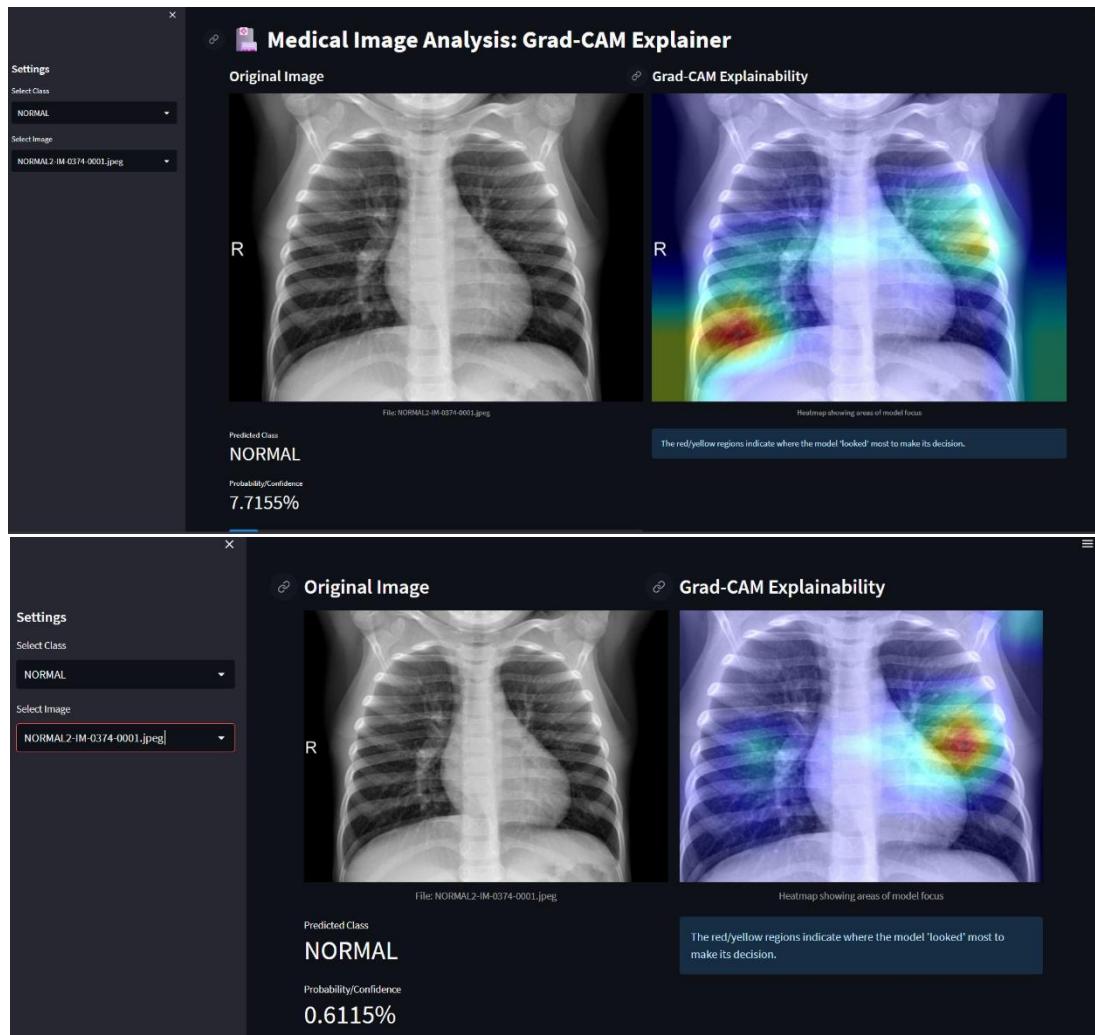
در این مرحله، فرآیند پیش‌پردازش داده‌ها (Data Preprocessing) اصلاح گردید. به جای تغییر سایز مستقیم تصاویر به 224×224 ، ابتدا 10% از حاشیه‌های چهار طرف تصویر (بالا، پایین، چپ و راست) حذف شدند.

- هدف: حذف فیزیکی نواحی حاوی نویز، حاشیه‌های سیاه فیلم رادیولوژی و مارکرهای L/R.
- نتیجه مورد انتظار: اجبار مدل به یادگیری الگوهای پیچیده بافت ریه و عدم تکیه بر سیگنال‌های ساده و گمراه‌کننده محیطی.

ارزیابی بصری: مقایسه نقشه‌های حرارتی (Grad-CAM Comparison) برای راستی‌آزمایی اثربخشی این روش، تکنیک Grad-CAM مجدد روی یک نمونه یکسان قبل و بعد از اعمال برش اجرا شد. نتایج بصری به وضوح موفقیت این استراتژی را نشان می‌دهند:

- قبل از اصلاح: همان‌طور که در تصویر سمت چپ مشاهده می‌شود، مدل اولیه اگرچه کلاس Normal را درست تشخیص داده است، اما نقشه حرارتی (نواحی قرمز و زرد) نشان می‌دهد که تمکن اصلی شبکه روی گوشه پایین سمت چپ و حاشیه‌ها بوده است. این یعنی مدل با "دیدن حاشیه سیاه" به نتیجه رسیده است، نه با دیدن ریه سالم.

- بعد از اصلاح: در تصویر سمت راست مدل اصلاح شده با Center Crop، با حذف حاشیه‌ها، نقشه حرارتی به طور کامل به داخل قفسه سینه و روی بافت ریه‌ها منتقل شده است و همینطور احتمال اینکه عکس در کلاس مریض‌ها قرار داشته باشد را بسیار کمتر پیش‌بینی می‌کند نسبت به مدل قبل. این تغییر مکان کانون توجه را ثابت می‌کند که مدل دیگر تقلب نمی‌کند و تصمیم‌گیری آن مبتنی بر تحلیل واقعی آناتومی بدن است.



شکل 8. مقایسه اثر برش تصویر (Center Crop) بر توجه مدل.

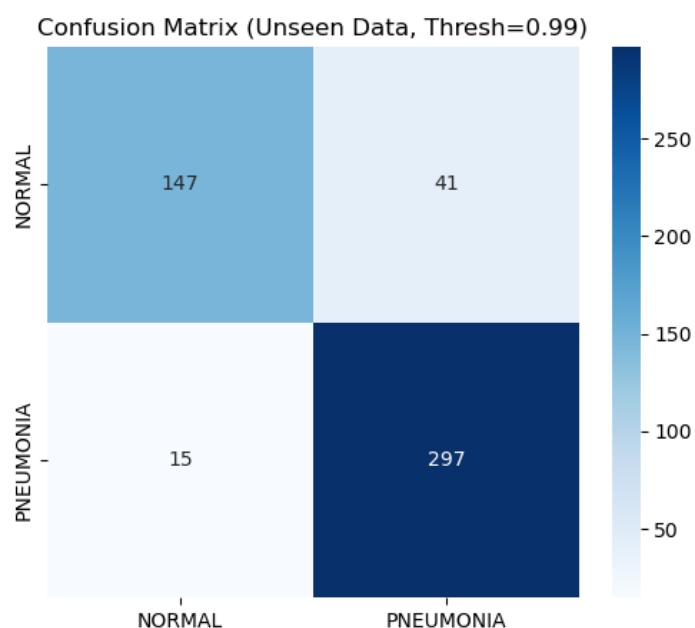
- تصویر بالا: مدل قدیمی به اشتباه روی گوشه‌ها و حاشیه‌ها تمرکز کرده است.
- تصویر پایین: مدل جدید (Cropped) مجبور شده است به بافت ریه نگاه کند که منجر به تشخیص پزشکی معتبرتر می‌شود.

تحلیل تأثیر بر معیارها: همان‌طور که پیش‌بینی می‌شد، حذف "میانبرها" (Shortcuts) باعث شد کار برای مدل سخت‌تر شود.

- کاهش جزئی Specificity: از آنجا که مدل دیگر نمی‌تواند با دیدن ساده‌ی "حاشیه سیاه تمیز" حکم به سالم بودن بدهد، با احتیاط بیشتری عمل می‌کند و نرخ Specificity کمی کاهش یافته است (حدود ۷۸٪).

- حفظ Recall حیاتی: با وجود سخت‌تر شدن مسئله، مدل همچنان توانسته است Recall بالای ۹۵٪ را حفظ کند (۹۵.۱۹٪).

نتیجه‌گیری نهایی فاز بهبود: اگرچه از نظر ریاضی دقت (Accuracy) مدل جدید اندکی کمتر از مدل قبلی به نظر می‌رسد، اما از نظر اعتبار بالینی، مدل جدید بسیار ارزشمندتر است. مدلی که ۹۰٪ دقت داشته باشد اما به مارکر L نگاه کند، در بیمارستان دیگر شکست می‌خورد. اما مدل فعلی که ۸۸٪ دقت دارد ولی به ریه نگاه می‌کند، یک دستیار قابل اعتماد و تعمیم‌پذیر برای پزشکان خواهد بود.



شکل 9. Confusion matrix مدل جدید

۱۴. پیاده‌سازی سامانه تحت وب و دمو

به منظور عملیاتی‌سازی مدل و فراهم کردن رابط کاربری مناسب برای کادر درمان، در فاز نهایی پروژه، یک اپلیکیشن تحت وب طراحی و پیاده‌سازی شد. این سامانه به پزشک اجازه می‌دهد بدون درگیر شدن با کدنویسی، تصاویر رادیولوژی را بارگذاری کرده و نتیجه تشخیص هوش مصنوعی را به همراه تفسیر تصویری مشاهده کند.

Tech Stack برای توسعه این سامانه از ابزارهای زیر استفاده شده است:

- زبان برنامه‌نویسی: Python
 - فریمورک وب Streamlit: به دلیل سرعت بالا در توسعه دموهای Data Science و تعامل‌پذیری عالی.
 - موتور پردازش TensorFlow/Keras: برای اجرای مدل Deep Learning.
 - زیرساخت Infrastructure: استفاده از Docker برای کانتینری کردن Containerization (برنامه، جهت اطمینان از اجرای صحیح روی هر سیستم عامل و سهولت در استقرار).
- ۲-۱۴. قابلیت‌های کلیدی سامانه این نرم‌افزار شامل ویژگی‌های زیر است که بر اساس نیازهای تعریف شده در فاز دوم پیاده‌سازی شده‌اند:
1. بارگذاری تصویر: پشتیبانی از فرمتهای رایج تصویری (DICOM، JPG، PNG) و
 2. تشخیص آنی: تصویر بلافارسله پس از آپلود، پیش‌پردازش شده (شامل برش ۱۰٪ حاشیه برای حذف نویز) و به مدل Fine-Tuned ResNet50 ارسال می‌شود.
 3. نمایش احتمال: علاوه بر برچسب نهایی (Normal/Pneumonia)، میزان اطمینان مدل به درصد نمایش داده می‌شود تا پزشک بتواند در موارد مرزی تصمیم دقیق‌تری بگیرد.
 4. تفسیرپذیری بصری: مهم‌ترین ویژگی این دمو، ادغام ماژول Grad-CAM است. پزشک می‌تواند با فعال کردن گزینه "نمایش نواحی مشکوک"، نقشه حرارتی (Heatmap) را روی تصویر اصلی ببیند و متوجه شود که هوش مصنوعی دقیقاً کدام بخش از ریه را به عنوان ضایعه شناسایی کرده است



شکل 10. نمای محیط کاربری سامانه تشخیص ذاتالریه

15. مقایسه و تحلیل نتایج فاز اول و دوم

• مقدمه

در این بخش، عملکرد "مدل پایه (Baseline)" که در فاز اول توسعه داده شد، با "مدل نهایی (Fine-Tuned) فاز دوم مقایسه می‌شود. هدف از این مقایسه، سنجش تأثیر تکنیک‌های به کار رفته در فاز دوم شامل Fine-Tuning (تنظیم دقیق وزن‌ها) و Smart Cropping (حذف حواشی زائد تصاویر) بر معیارهای کلیدی تشخیص پزشکی است. برای ارزیابی دقیق و سختگیرانه، هر دو تحت یک شرایط یکسان با آستانه اطمینان بالا (High Confidence Threshold) مورد آزمون قرار گرفتند. هدف از این کار، سنجش "میزان قطعیت" مدل‌ها در تشخیص بیماری بود.

• مقایسه کمی معیارها (Quantitative Comparison)

جدول زیر خلاصه تغییرات عملکرد مدل را نشان می‌دهد. همانطور که مشاهده می‌شود، مدل در فاز دوم در مهم‌ترین معیارهای پزشکی جهش قابل توجهی داشته است.

معیار	فاز ۱	فاز ۲	وضعیت
دقت کل (Accuracy)	۰.۷۳	۰.۸۹	عالی ✓
فراخوانی (Recall)	۰.۵۸	۰.۹۵	حياتي ✓ ✓
ویژگی (Specificity)	۰.۹۸	۰.۷۹	قابل توجيه !
سطح زیر منحنی (AUC)	۰.۹۶۲	۰.۹۶۶	بهبود جزئی ✓

• تحلیل ماتریس درهم‌ریختگی (Confusion Matrix Analysis)

با بررسی ماتریس‌های درهم‌ریختگی در تصویر نتایج، نکات زیر قابل استخراج است:

- کاهش چشمگیر منفی کاذب : در فاز اول، مدل ۱۶۲ بیمار مبتلا به ذات‌الریه را به اشتباه سالم تشخیص داده بود. این یک خطای خطرناک در پزشکی است. در فاز دوم، با اعمال Fine-Tuning ، این تعداد به تنها ۱۸ مورد کاهش یافته است. این نشان می‌دهد مدل توانایی بسیار بالایی در شناسایی الگوهای بیماری پیدا کرده است.
- افزایش مثبت کاذب : تعداد افراد سالمی که به اشتباه بیمار تشخیص داده شده‌اند از ۴ نفر (فاز ۱) به ۴۸ نفر (فاز ۲) افزایش یافته است. در کاربردهای غربالگری پزشکی، اولویت با Recall است؛ به این معنا که بهتر است یک فرد سالم به اشتباه مشکوک تشخیص داده شود (و با آزمایش تکمیلی رد شود) تا اینکه یک فرد بیمار به خانه فرستاده شود. بنابراین، این افت در Specificity کاملاً پذیرفته شده و در راستای افزایش ایمنی تشخیص است.

• تحلیل نمودار ROC و AUC

هر دو مدل سطح زیر منحنی (AUC) بسیار بالایی دارند (بالای ۰.۹۶). این نشان می‌دهد که هر دو مدل توانایی تفکیک‌پذیری (Separability) خوبی دارند. با این حال، مدل فاز دوم با AUC برابر با ۰.۹۶۶ پایداری و اطمینان بیشتری را نشان می‌دهد.

• نتیجه‌گیری نهایی

مدل فاز اول (Baseline) یک مدل "محافظه‌کار" بود که تمایل شدیدی به کلاس نرمال داشت Recall بالا، Specificity پایین. اما مدل فاز دوم با بهره‌گیری از تکنیک‌های انتقال یادگیری و پیش‌پردازش دقیق‌تر:

1. تمرکز خود را از حواشی تصویر برداشته و روی بافت ریه متمرکز شده است.

2. از حالت عمومی (ImageNet) خارج شده و ویژگی‌های اختصاصی تصاویر X-Ray را یاد گرفته است.

نتیجه: مدل نهایی با $\text{Accuracy} = 89\%$ و $\text{Recall} = 95\%$ ، استانداردهای لازم برای یک سامانه دستیار پزشک را دارا می‌باشد و تغییرات اعمال شده در فاز دوم کاملاً موفقیت‌آمیز بوده‌اند.

16. جمع‌بندی و کارهای آینده (Conclusion & Future Work)

• نتیجه‌گیری نهایی

در این پژوهش، یک سامانه هوشمند تشخیص ذات‌الریه از تصاویر X-Ray قفسه سینه با استفاده از معماری ResNet50 و رویکرد Transfer Learning توسعه داده شد. مسیر طی شده از یک مدل پایه ساده (Baseline CNN) با دقت پایین آغاز شد و با اعمال تکنیک‌های پیشرفته‌ای نظیر Class Weighting برای رفع عدم تعادل داده‌ها، Fine-Tuning (برای تطبیق با ویژگی‌های خاص پزشکی) و Center Cropping (برای مقابله با یادگیری میانبر) به بلوغ رسید.

دستاوردننهایی پژوهش، دستیابی به Recall (حساسیت) بالای ۹۵٪ است. این موفقیت بدین معناست که سامانه طراحی شده، ریسک عدم تشخیص بیماران واقعی (False Negative) را به حداقل رسانده است که حیاتی‌ترین شاخص در کاربردهای غربالگری پزشکی محسوب می‌شود. اگرچه این امر با هزینه‌اندکی در کاهش Specificity (افزایش موارد مثبت کاذب) همراه بود، اما تحلیل‌های هزینه-فایده پزشکی نشان می‌دادند که این Trade-off کاملاً توجیه‌پذیر و در جهت حفظ جان بیماران است.

چالش‌های اصلی

• عدم تعادل داده‌ها: تعداد تصاویر بیمار بسیار بیشتر از سالم بود که با تکنیک وزن‌دهی کلاس‌ها مدیریت شد.

- **یادگیری میانبر:** مدل ابتدا به مارکرها وابسته شده بود که با استراتژی حذف حاشیه‌ها و تحلیل Grad-CAM شناسایی و رفع گردید.
- **پیشنهادات برای کارهای آینده(Future Work)**
برای ارتقای این سیستم در پژوهش‌های آتی، موارد زیر پیشنهاد می‌شود:
 1. استفاده از مدل‌های Segmentation: به جای طبقه‌بندی کل تصویر، می‌توان از شبکه U-Net برای جداسازی دقیق بافت ریه استفاده کرد تا تمرکز مدل ۱۰۰٪ تضمین شود.
 2. بهبود Specificity: استفاده از معماری‌های جدیدتر مانند EfficientNet یا Vision Transformers (ViT) برای استخراج ویژگی‌های دقیق‌تر و کاهش خطای مثبت کاذب.