

به نام خدا

پروژه پایانترم درس هوش مصنوعی

اعضا: دانیال خاکساری – سید طاها طیبی – امیرعلی یاوری

موضوع پروژه : تشخیص بیماری ذات الریه با استفاده از عکس های X-Ray

تاریخ:زمستان 1404

لینک گیت هاب پروژه:

[YavariAmirali/AI-FinalProject-Tayebi: Medical Imaging AI Project - Pneumonia Detection using Chest X-Ray](https://github.com/YavariAmirali/AI-FinalProject-Tayebi)

لیست محتوا

[1.تعریف دقیق مسئله (Problem Definition) 3](#_Toc221030655)

[2.مهم‌ترین اهداف و معیارهای پروژه 3](#_Toc221030656)

[3.توجیه انتخاب معماری و استراتژی مدل‌سازی (Model Justification) 4](#_Toc221030657)

[فاز اول: مدل پایه (Simple CNN Baseline) 4](#_Toc221030658)

[فاز دوم: یادگیری انتقالی با ResNet50 (Transfer Learning) 4](#_Toc221030659)

[دلایل انتخاب ResNet50: 5](#_Toc221030660)

[4.تحلیل اکتشافی داده ها: (EDA) 5](#_Toc221030661)

[5.جدول پلن ازمایش ها 9](#_Toc221030662)

[6.خروجی ساختار مدل (Model Summary) 10](#_Toc221030663)

[7.معرفی دیتاست: 11](#_Toc221030664)

[8.جمع‌بندی: 11](#_Toc221030665)

# فاز یک

## 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.

## مهم‌ترین اهداف و معیارهای پروژه

Recall (Sensitivity) بالاتر از ۹۵٪:

Recall 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.

Accuracy بالاتر از ۹۰٪:

این معیار به منظور سنجش عملکرد کلی مدل در تشخیص صحیح تصاویر سالم و بیمار مورد استفاده قرار می‌گیرد. اگرچه Accuracy به تنهایی معیار قابل اتکایی 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.

## 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 دچار 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 بودs

### فاز دوم: یادگیری انتقالی با ResNet50 (Transfer Learning)

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

دلایل انتخاب ResNet50:

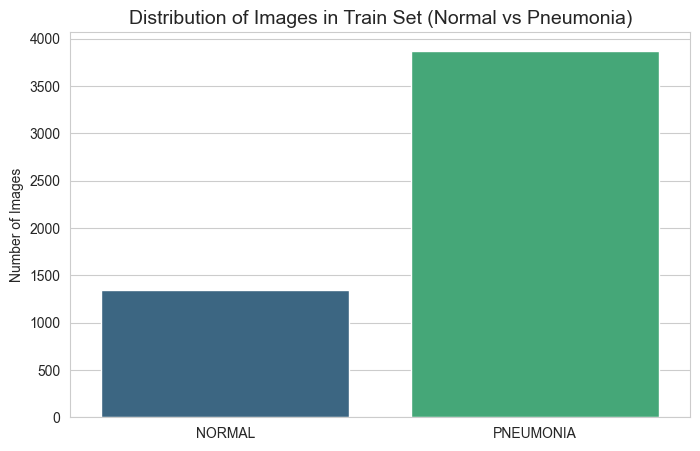
\* استفاده از Residual Connections که مشکل Vanishing Gradient را در شبکه‌های عمیق کاهش می‌دهد

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

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

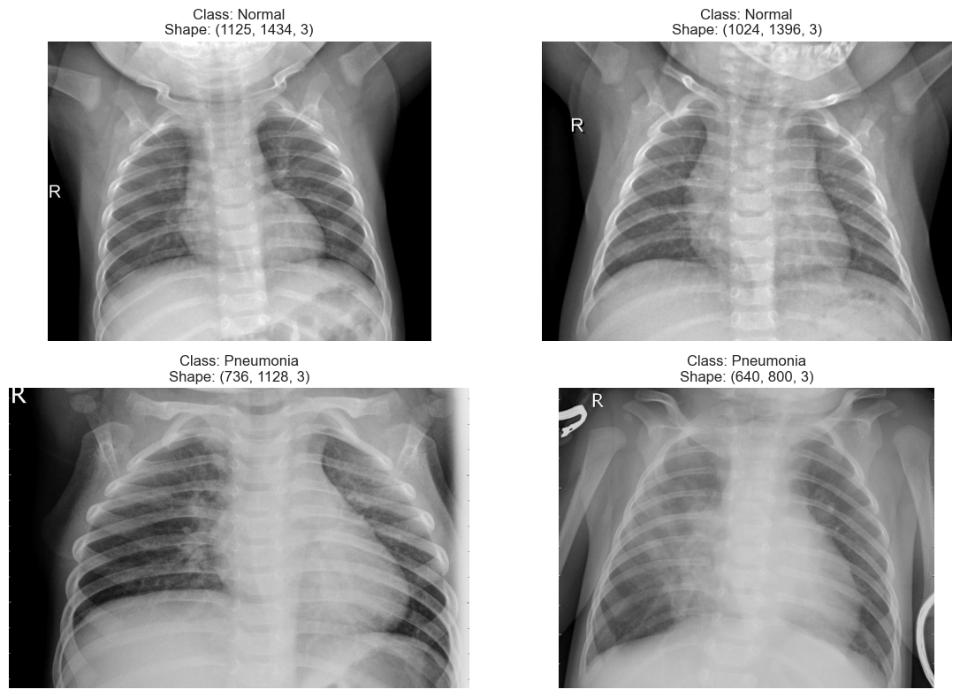
در این فاز، لایه‌های ابتدایی شبکه که ویژگی‌های عمومی بصری مانند لبه‌ها و بافت‌ها را استخراج می‌کنند، فریز (Freeze) شده و فرآیند Fine-tuning عمدتاً بر روی لایه‌های انتهایی انجام می‌گیرد تا مدل به‌طور مؤثر با داده‌های پزشکی X-Ray تطبیق یابد.

## تحلیل اکتشافی داده ها: (EDA)



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**Class Weighting**برای رفع این مشکل استفاده خواهیم کرد.



شکل1. 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 کنتراست و روشنایی تصاویر دیده‌ می‌شود که لزوم نرمال‌سازی را تایید می‌کند.



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

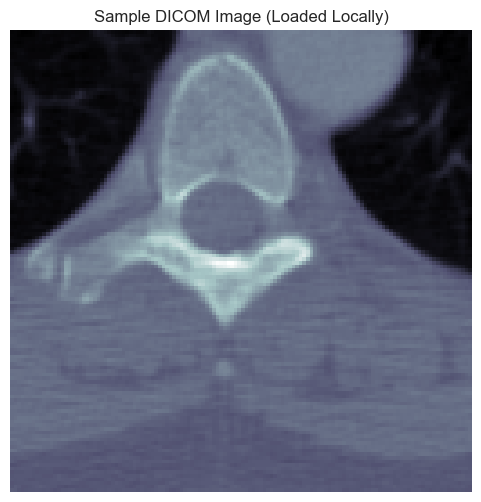
این نمودار نشان می‌دهد که تصاویر دیتاست دارای رزولوشن‌های متنوعی هستند و سایز یکسانی ندارند. از آنجا که معماری‌های استاندارد CNN مانند مدلی که طراحی خواهیم کرد و ResNet50 نیاز به ورودی با ابعاد ثابت دارند، تمامی تصاویر در مرحله پیش‌پردازش به سایز استاندارد **224\*224پیکسل** تغییر سایز (Resize) خواهند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

A graph with purple lines

Description automatically generated

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 (Scale) خواهیم کرد.



شکل 2.نمونه ای از عکس های DICOM

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

## جدول پلن ازمایش ها

| ویژگی (Feature) | مدل فاز اول (Baseline) | مدل فاز دوم (Final Phase) |
| --- | --- | --- |
| نام مدل | Simple CNN (دست‌ساز) | ResNet50 (Transfer Learning) |
| معماری | ۳ لایه کانولوشن ساده + MaxPooling | ۵۰ لایه عمیق با وزن‌های ImageNet |
| ورودی داده | تغییر سایز (Resize) و نرمال‌سازی ساده | پیش‌پردازش  ImageNet + Data Augmentation |
| مدیریت عدم تعادل | ندارد (None) | Class Weighting (وزن‌دهی کلاس‌ها) |
| هدف آزمایش | بررسی امکان‌پذیری (Feasibility) و ایجاد خط مبنا | دستیابی به Recall > 95% و قابلیت اطمینان بالا |
| معیارهای ارزیابی | Accuracy, Recall | Accuracy, Recall, Specificity, AUC |

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

دلیل پزشکی: در تشخیص بیماری‌ها، فقط پیدا کردن بیماران (Recall) کافی نیست؛ بلکه باید مطمئن شویم افراد سالم به اشتباه بیمار تشخیص داده نمی‌شوند (False Positive). تشخیص اشتباه یک فرد سالم به عنوان بیمار، می‌تواند هزینه‌های روانی سنگین و درمان‌های غیرضروری را به بیمار تحمیل کند. بنابراین معیار Specificity در کنار Recall برای ما حیاتی است

## 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  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## معرفی دیتاست:

برای آموزش و ارزیابی مدل هوش مصنوعی در این پروژه، از مجموعه داده معتبر Chest X-Ray Images (Pneumonia) 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 (مبتلا به ذات‌الریه) طبقه‌بندی شده‌اند. تصاویر متعلق به بیماران اطفال 1 تا 5 سال بوده و در مرکز پزشکی زنان و کودکان گوانگژو جمع‌آوری شده‌اند.

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

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

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

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

لینک دیتاست: [Chest X-Ray Images (Pneumonia)](https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia?resource=download)

## جمع‌بندی:

استراتژی دو‌مرحله‌ای اتخاذشده 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 اتکا 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

## 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 ImageNet آموزش دیده است.

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

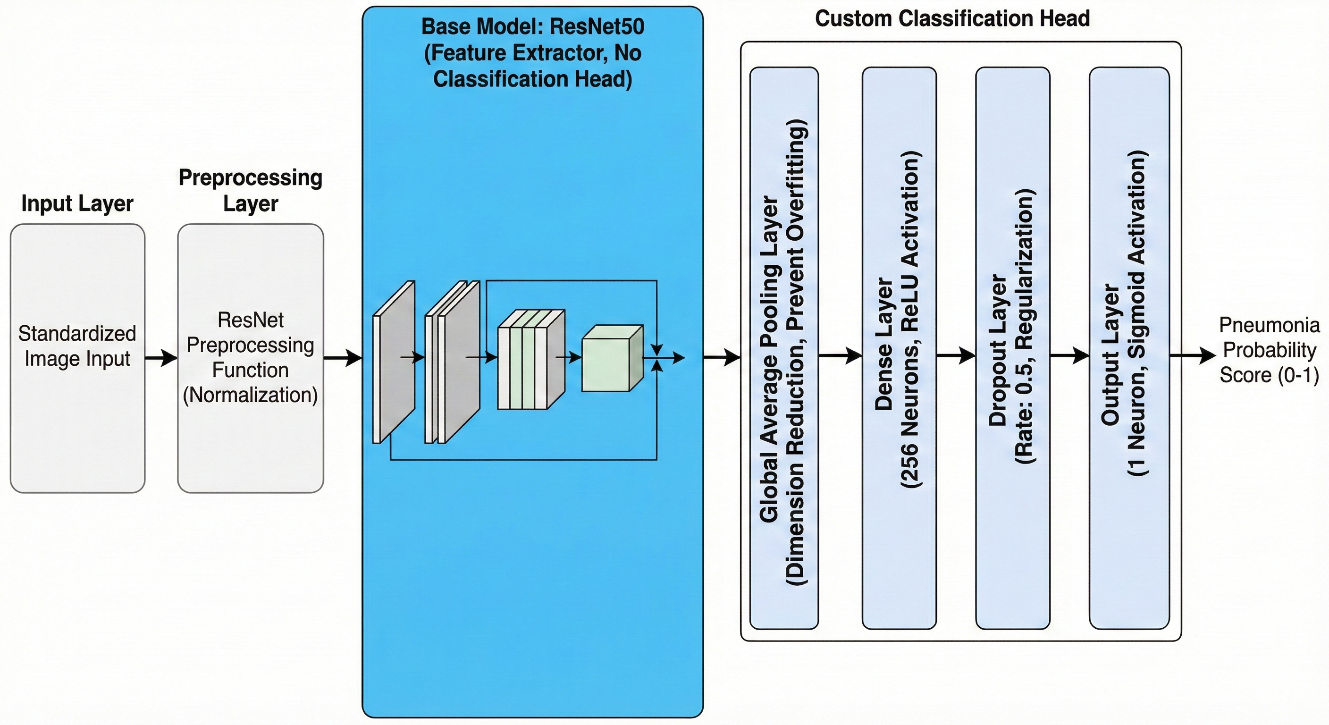
لایه ورودی: دریافت تصاویر با ابعاد استاندارد شده.

پیش‌پردازش(Preprocessing): اعمال تابع پیش‌پردازش مخصوص ResNet برای نرمال‌سازی 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:

* Global Average Pooling: برای کاهش ابعاد ویژگی‌ها و جلوگیری از Overfitting.
* Dense Layer (256 Neurons): با تابع فعال‌ساز ReLU برای یادگیری ویژگی‌های غیرخطی.
* Dropout (0.5): خاموش کردن تصادفی ۵۰٪ نورون‌ها در حین آموزش برای افزایش تعمیم‌پذیری مدل.
* Output Layer (1 Neuron): لایه خروجی با تابع فعال‌ساز Sigmoid 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 لایه‌های طبقه‎بندی (Feature Extraction) در ابتدا، تمامی لایه‌های ResNet50 در حالت فریز (Frozen) قرار گرفتند .(trainable=False) در این مرحله، تنها وزن‌های لایه‌های 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 آموزش‌ندیده.

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 مجدد انتخاب شدند.

تنظیمات حیاتی: در این مرحله، مدل با نرخ یادگیری بسیار پایین کامپایل 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.

## نتایج آموزش و ارزیابی (Training & Evaluation Results)

### روند آموزش مدل (Training History)

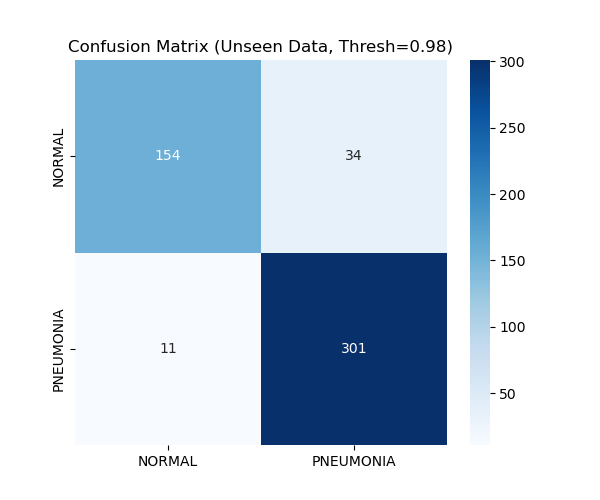
در فاز دوم، مدل Fine-Tuned با نرخ یادگیری بسیار پایین و با استفاده از بهینه‌ساز **Adam** آموزش داده شد. برای جلوگیری از Overfitting و حفظ تعمیم‌پذیری مدل روی داده‌های جدید، از تکنیک **Early Stopping** با پایش خطای اعتبارسنجی (val\_loss) استفاده گردید.

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 درهم‌ریختگی 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 اکثریت 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 پزشکی 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 تقریبا 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 گزارش‌شده 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) | **حیاتی‌ترین 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. |

## 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.

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

**جدول مقایسه عملکرد (Threshold = 0.98) :**

**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 درRecall (+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 فوق‌العاده بالا، ما مجبور شدیم آستانه را طوری تنظیم کنیم که مدل "محتاط‌تر" عمل کند. این موضوع باعث شد معیار Specificity حدود ۴.۸٪ کاهش یابد و تعداد (False Positives) از ۲۵ به 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.

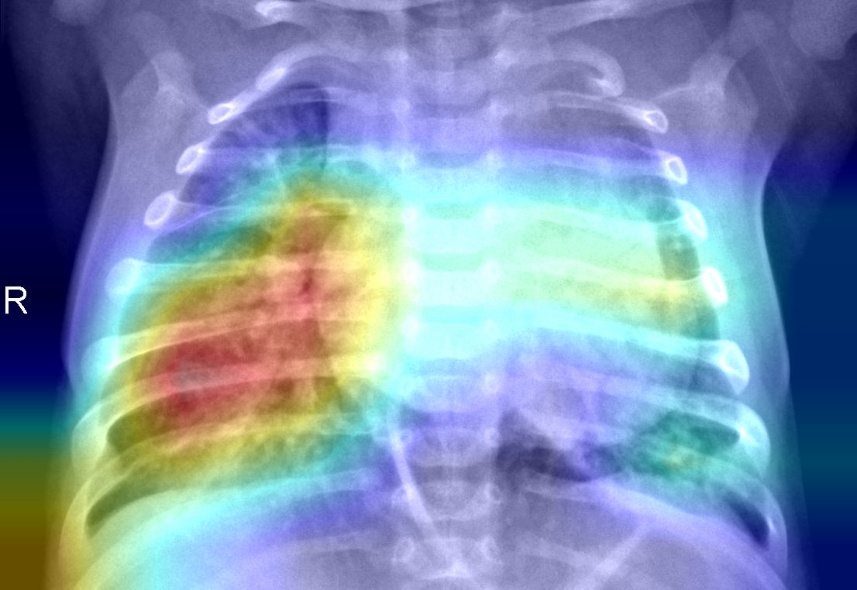
**نتیجه‌گیری:** مدل نهایی با دستیابی به **Recall > 96%** و **Accuracy > 91%**، تمامی اهداف تعیین شده در پروپوزال پروژه را با موفقیت پشت سر گذاشته است.

## 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 پروژه 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 5.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 خطا و یادگیری میانبر (Shortcut Learning Analysis)

با بررسی موارد خطا 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 خطا :

شناسایی این خطا 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 تشخیص‌دهنده.

## پیاده‌سازی و ارزیابی راهکار حذف یادگیری میانبر

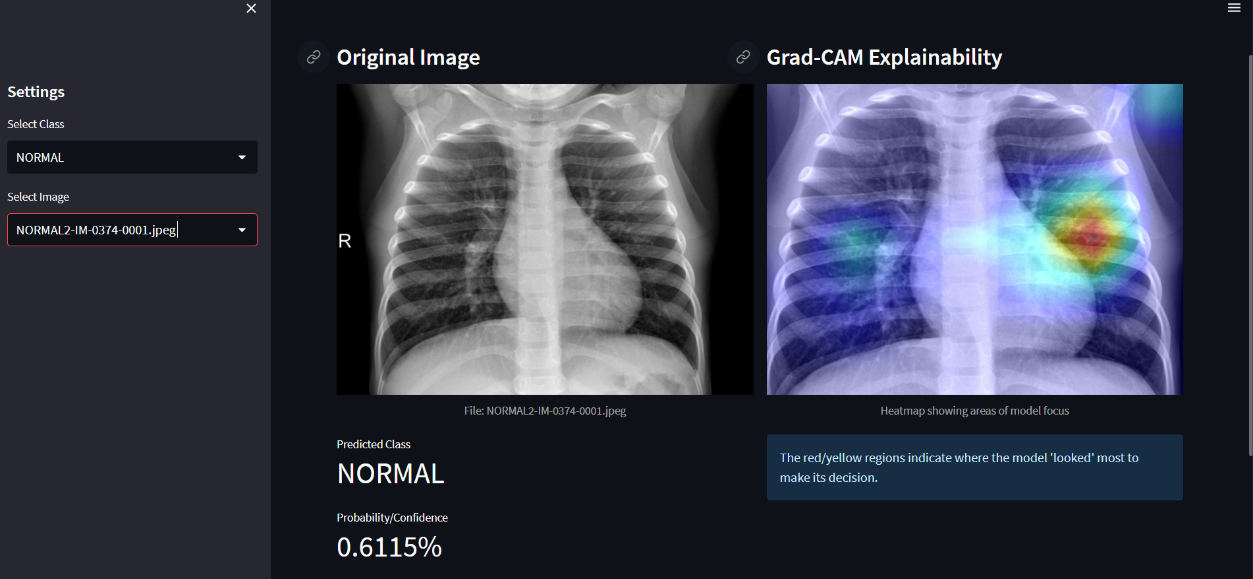
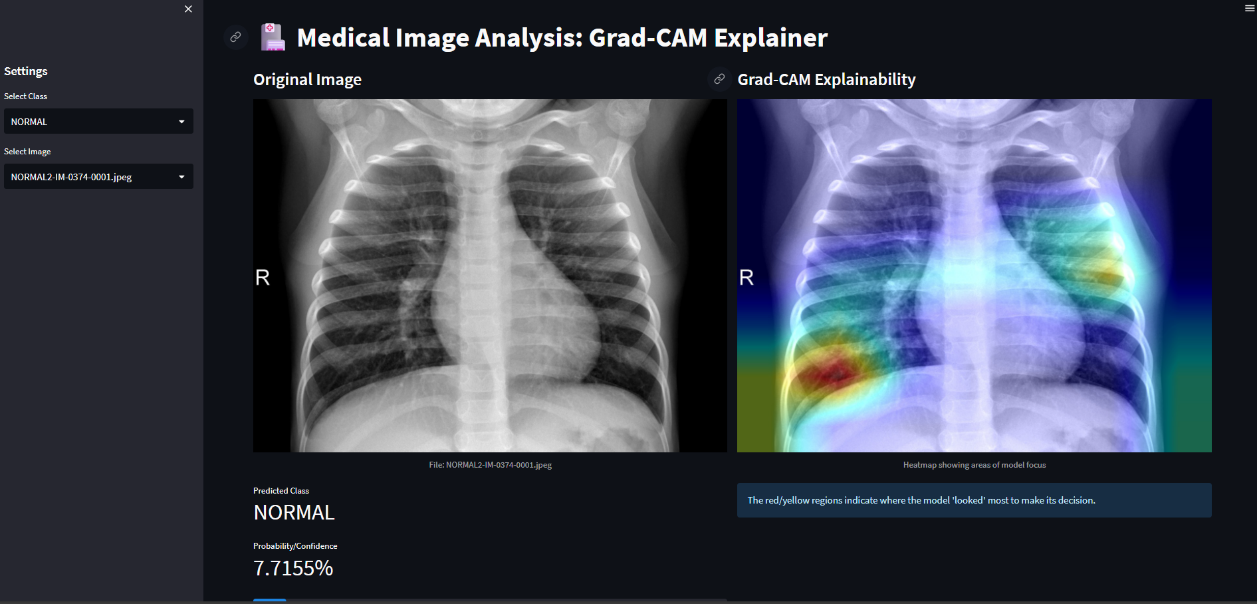
پیرو تحلیل‌های انجام شده در بخش قبل و شناسایی پدیده Shortcut Learning، که در آن مدل به جای بافت ریه به حاشیه‌های تصویر و مارکرها توجه می‌کرد، راهکار **برش مرکز** به عنوان مؤثرترین روش برای رفع این بایاس پیاده‌سازی شد.

در این مرحله، فرآیند پیش‌پردازش داده‌ها (Data Preprocessing) اصلاح گردید. به جای تغییر سایز مستقیم تصاویر به ۲۲۴×۲۲۴، ابتدا **۱۰٪ 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 مجدداً 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 تمرکز 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 باCenter 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 مریض ها قرار داشته باشد را بسیار کمتر پیش بینی 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 جزئیSpecificity:** 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 (Generalizable) برای پزشکان خواهد بود.