

گزارش کار - تشخیص سرطان ریه با شبکه‌های عصبی عمیق - ملیکا بلبل آبادی

در این پروژه، یک مدل deep learning برای تشخیص سرطان ریه از تصاویر اشعه‌ی ایکس توسعه داده شد. ابتدا یک شبکه‌ی CNN طراحی کردم و سپس با مدل از پیش‌آموزش‌یافته‌ی ResNet50 مقایسه کردم. هدف از این کار، بررسی دقت مدل‌های مختلف و بهینه‌سازی عملکرد آنها برای تشخیص صحیح سه کلاس مختلف بود:

- Benign cases (خوش‌خیم)
- Malignant cases (بدخیم)
- Normal cases (نرمال)

- داده‌ها و پیش‌پردازش

داده‌های این پروژه از مجموعه‌ی IQ-OTHNCCD Lung Cancer Dataset (نسخه‌ی Augmented) دریافت شد که شامل تصاویر اشعه‌ی ایکس بیماران مبتلا به سرطان خوش‌خیم، بدخیم و همچنین افراد سالم بود.

- پیش‌پردازش داده‌ها

مراحل پردازش داده‌ها شامل:

- تغییر اندازه‌ی تصاویر به 224×224 پیکسل
- انجام Data Augmentation برای افزایش تنوع داده‌ها:
- چرخش تصادفی (RandomRotation)
- وارونه‌سازی افقی (RandomHorizontalFlip)
- افزایش تنوع رنگی و کنتراست (ColorJitter)
- حذف تصادفی بخش‌هایی از تصویر (RandomErasing)
- نرمال‌سازی مقادیر پیکسلی به $[0.5, 0.5, 0.5]$
- تقسیم داده‌ها به ۸۰٪ آموزش و ۲۰٪ اعتبارسنجی

مدل اول: طراحی CNN از ابتدا

این مدل شامل ۴ لایه کانولوشنی همراه با نرمال‌سازی، توابع فعال‌سازی ReLU و لایه‌های Dropout برای کاهش بیش‌برازش بود. همچنین برای بهینه‌سازی، از Adam به همراه کاهش نرخ یادگیری استفاده شد.

ساختار مدل

- ۴ لایه‌ی کانولوشنی همراه با Batch Normalization و MaxPooling
- استفاده از Global Average Pooling به جای Flatten برای کاهش تعداد پارامترها
- لایه‌های Fully Connected همراه با Dropout برای کاهش بیش‌برازش

تنظیمات آموزش مدل

پارامترهای آموزش:

- تابع هزینه: CrossEntropyLoss همراه با Label Smoothing و Class Weights
- بهینه‌ساز: Adam با $lr=0.0002$ و $weight_decay=0.001$
- استراتژی کاهش نرخ یادگیری: ReduceLROnPlateau
- تعداد اپاک‌ها: ۳۰ اپاک
- استفاده از Early Stopping برای جلوگیری از بیش‌برازش

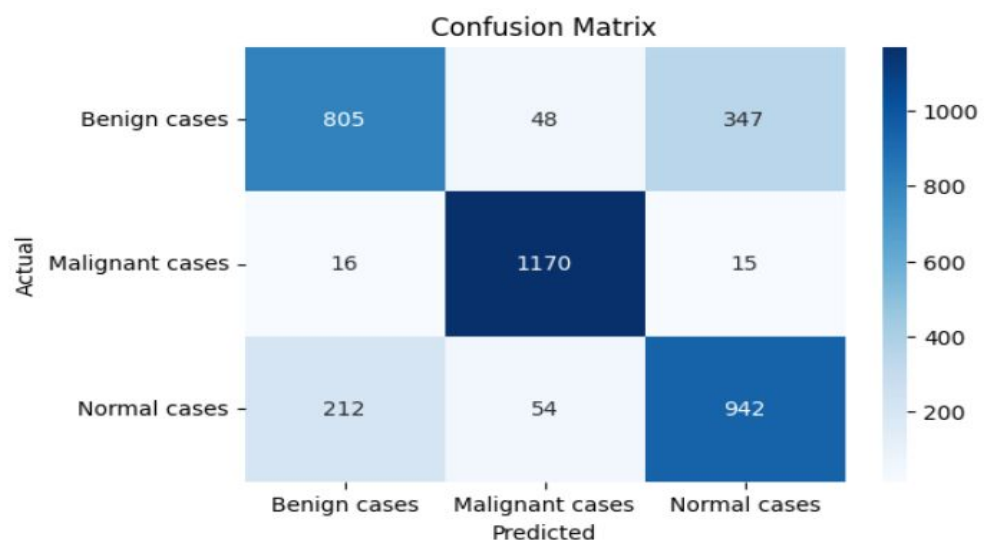
نتایج نهایی CNN

- Train Accuracy: 80.74%
- Validation Accuracy: 83.32%
- Precision, Recall, F1-score

Benign cases → Precision: 79% | Recall: 73% | F1-score: 76%
Malignant cases → Precision: 93% | Recall: 98% | F1-score: 95%
Normal cases → Precision: 76% | Recall: 78% | F1-score: 77%

مشاهده‌ی رشد مدل:

- نوسانات در اپاک‌های اولیه وجود داشت ولی پس از چند اپاک، مدل به دقت بالای ۸۰٪ رسید.
- کلاس سرطان بدخیم (Malignant) بهترین عملکرد را داشت (F1-score: 95%).



Validation Loss: 0.6274, Validation Accuracy: 80.58%

Validation Loss: 0.6248, Validation Accuracy: 81.27%

Validation Loss: 0.6215, Validation Accuracy: 81.02%

Validation Loss: 0.6247, Validation Accuracy: 81.19%

Validation Loss: 0.6205, Validation Accuracy: 81.91%

Validation Loss: 0.6271, Validation Accuracy: 80.66%

Epoch [27/30] → Loss: 0.6122, Accuracy: 78.17%

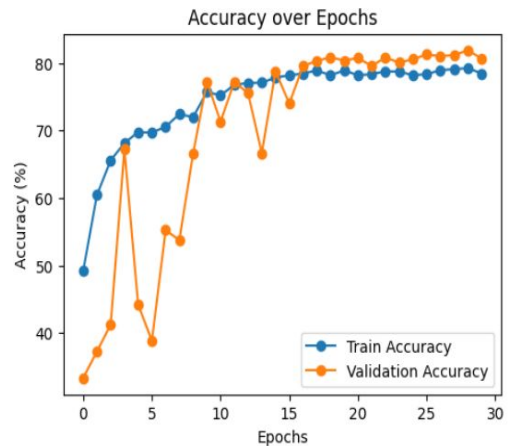
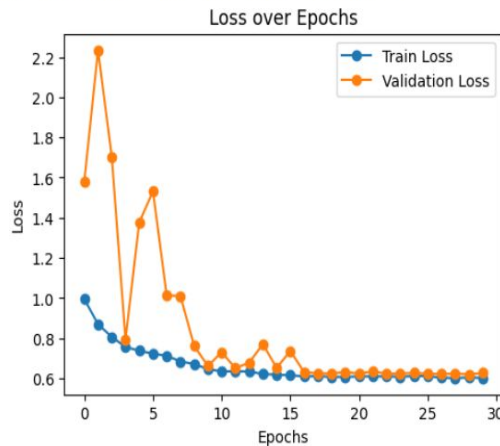
Epoch [26/30] → Loss: 0.6121, Accuracy: 78.36%

Epoch [27/30] → Loss: 0.6038, Accuracy: 78.94%

Epoch [28/30] → Loss: 0.6023, Accuracy: 79.11%

Epoch [29/30] → Loss: 0.6050, Accuracy: 79.25%

Epoch [30/30] → Loss: 0.6031, Accuracy: 78.47%



مدل دوم: استفاده از ResNet50 پیش‌آموزش‌یافته

این مدل یک شبکه از پیش آموزش‌دیده روی ImageNet است که برای داده‌های ما Fine-tune شد. لایه‌های اولیه آن ثابت نگه داشته شدند و تنها لایه‌ی Fully Connected آن برای خروجی 3 کلاس تغییر کرد. این مدل دارای تعمیم‌پذیری بالا و توانایی یادگیری ویژگی‌های عمیق‌تر است.

تنظیمات مدل ResNet50

- استفاده از مدل ResNet50 با وزن‌های پیش‌آموزش‌یافته‌ی ImageNet
- فریز کردن لایه‌های اولیه و فقط آموزش fc برای تشخیص سه کلاس
- تغییر تعداد خروجی‌ها به ۳ کلاس

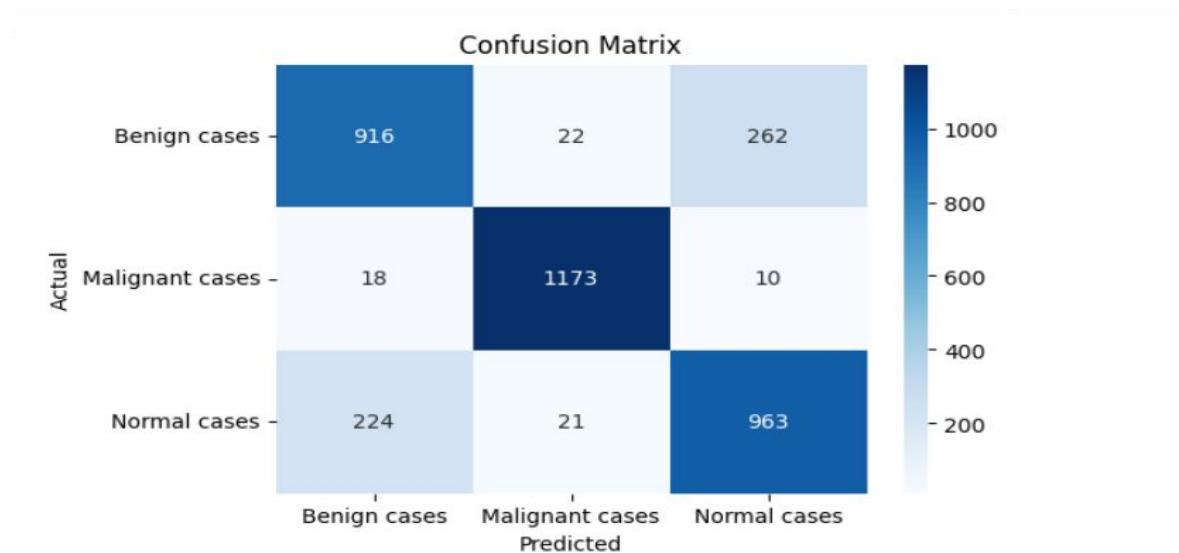
تنظیمات آموزش مدل

پارامترهای آموزش:

- تابع هزینه: CrossEntropyLoss
- بهینه‌ساز: Adam با lr=0.001 و weight_decay=0.001
- استراتژی کاهش نرخ یادگیری: ReduceLROnPlateau
- تعداد اپاک‌ها: ۳۰ اپاک

نتایج نهایی ResNet50

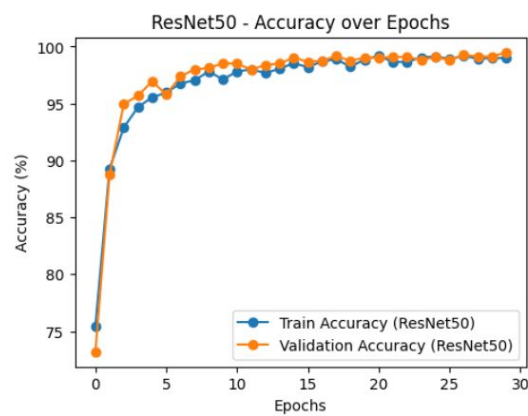
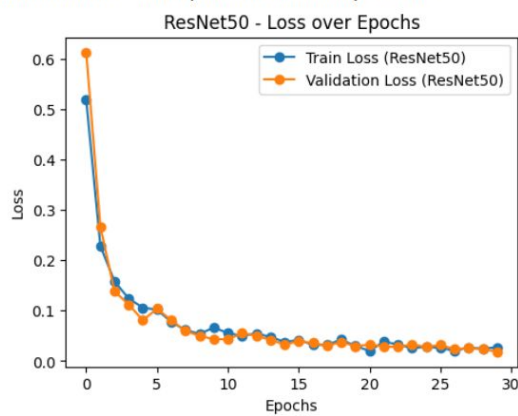
- بهترین دقت روی مجموعه‌ی اعتبارسنجی: 99.50%
- دقت فوق‌العاده بالا در همه‌ی کلاس‌ها
- F1-score کلی: 0.99
- همگرایی سریع‌تر نسبت به CNN



Epoch [27/30] → Train Loss: 0.0255, Train Accuracy: 99.17%
 Validation Loss: 0.0235, Validation Accuracy: 99.25%
 Epoch [28/30] → Train Loss: 0.0263, Train Accuracy: 98.95%
 Validation Loss: 0.0257, Validation Accuracy: 99.14%
 Epoch [29/30] → Train Loss: 0.0247, Train Accuracy: 98.97%
 Validation Loss: 0.0232, Validation Accuracy: 99.11%

Validation Loss: 0.0184, Validation Accuracy: 99.50%

Epoch [30/30] → Train Loss: 0.0270, Train Accuracy: 99.00%



مقایسه‌ی CNN و ResNet50

بررسی معیارهای زیر:

1. Accuracy (دقت کلی)
2. Confusion Matrix برای بررسی خطاهای مدل
3. Precision, Recall, F1-score برای تحلیل عملکرد در هر کلاس
4. سرعت و زمان آموزش هر مدل
5. تفاوت عملکرد روی داده‌های واقعی (تست با تصاویر جدید)

ResNet50	CNN سفارشی	معیار
99.50 %	83.32 %	بالاترین دقت اعتبارسنجی
0.83	0.99	F1-score
کمتر	بیشتر	نوسان در Loss
سریع‌تر	آهسته‌تر	سرعت همگرایی
عالی	خوب	توانایی تعمیم‌دهی
بسیار بالا	پایین	پیچیدگی مدل

تحلیل

• ResNet50 عملکرد بسیار بهتری نسبت به CNN سفارشی داشت و در کمتر از 10 اپیاک دقتش از 95% عبور کرد.

• CNN به حدود 30 اپیاک نیاز داشت تا به دقت 83% برسد، اما ResNet50 تنها در 3-4 اپیاک از این مقدار عبور کرد.

• بیش‌برازش (Overfitting) در ResNet50 مشاهده نشد، زیرا دقت روی داده‌های اعتبارسنجی تقریباً برابر با دقت روی داده‌های آموزشی باقی ماند.

• پیش‌آموزش روی ImageNet کمک کرد تا ویژگی‌های پیچیده‌تر و مهم‌تر را سریع‌تر استخراج کند.

تجیه‌گیری

• ResNet50 برنده‌ی این مقایسه است! این مدل دقت بسیار بالاتر، سرعت یادگیری سریع‌تر و تعمیم‌پذیری بهتری داشت.

• در کاربردهای واقعی مثل تشخیص پزشکی، دقت بالاتر اهمیت بسیار زیادی دارد و ResNet50 گزینه‌ی مناسب‌تری محسوب می‌شود.

• در صورتی که سخت‌افزار و منابع کافی در اختیار باشد، استفاده از مدل‌های پیش‌آموزش‌دیده مانند ResNet50 توصیه می‌شود.

• با این حال، مدل‌های سفارشی نیز می‌توانند در شرایطی که داده‌های بسیار خاص و محدود داریم، تنظیم‌پذیری بیشتری داشته باشند.

پیشنهادهای برای بهبود بیشتر

استفاده از Augmentation بیشتر مانند CutMix یا MixUp برای افزایش تعمیم‌پذیری.

آزمایش مدل‌های سبک‌تر مثل MobileNet یا EfficientNet برای بهینه‌سازی سرعت و مصرف منابع.

افزایش حجم داده‌ها یا استفاده از Transfer Learning از مدل‌هایی که روی داده‌های مشابه آموزش دیده‌اند.

چرا ResNet50 دچار Overfitting نشده؟

1. دقت روی مجموعه‌ی اعتبارسنجی به دقت مجموعه‌ی آموزش نزدیک است:

• دقت آموزش در پایان 99.00%

• دقت اعتبارسنجی در پایان 99.50%

• این نشان می‌دهد که مدل توانسته روی داده‌های جدید هم عملکرد مشابهی داشته باشد و

دچار بیش‌برازش نشده است.

2. عدم وجود شکاف زیاد بین Loss آموزش و اعتبارسنجی:

• اگر مدل دچار Overfitting شده بود، انتظار داشتیم که Loss اعتبارسنجی افزایش یابد و مدل

روی داده‌های جدید عملکرد ضعیف‌تری داشته باشد.

• اما در اینجا، Loss اعتبارسنجی در طول آموزش کاهش یافته است و حتی در ایپاک‌های آخر

هم کاهش خوبی دارد.

3. F1-score بالا در تمامی کلاس‌ها:

• اگر Overfitting اتفاق می‌افتاد، مدل ممکن بود فقط یک کلاس خاص را به درستی تشخیص

دهد و در کلاس‌های دیگر عملکرد بدی داشته باشد.

• اما اینجا هر سه کلاس Benign، Malignant و Normal با دقت بالایی تشخیص داده شده‌اند.

چطور مطمئن بشیم که در آینده دچار Overfitting نمیشیم؟

افزایش داده‌ها یا استفاده از Data Augmentation بیشتر مثل CutMix یا MixUp.

استفاده از Dropout در لایه Fully Connected (که در حال حاضر 0.5 در نظر گرفته شده است).

کاهش تعداد ایپاک‌ها در صورت مشاهده‌ی افزایش Loss اعتبارسنجی در ایپاک‌های پایانی.

استفاده از Early Stopping (که در اینجا انجام شد) برای جلوگیری از آموزش بیش از حد.