



دانشکده مهندسی برق

گروه مهندسی برق گرایش هوش مصنوعی و رباتیک

گزارش کارآموزی

تعیین سختی جراحی آب مروارید با استفاده از هوش مصنوعی

نام و نام خانوادگی دانشجو: یزدان بیات

نام و نام خانوادگی استاد کارآموزی: دکتر تقیراد

محل کارآموزی: آزمایشگاه رباتیک ارس

آدرس: دانشکده مهندسی برق دانشگاه خواجه نصیر

تاریخ انجام کارآموزی: تابستان 1404

الحمد لله الذي
خلقنا من
الحمم

تشکر و قدردانی

بدین وسیله مراتب قدردانی و سپاس صمیمانه خود را از استاد ارجمند جناب آقای دکتر تقیراد که با راهنمایی‌ها و دانش ارزشمند خود نقش مهمی در پیشبرد این دوره کارآموزی داشتند، ابراز می‌دارم. همچنین از تیم محترم ارس و تمامی بزرگوارانی که در طول این مسیر با حمایت و همراهی خود یاری‌رسان بودند، صادقانه تشکر می‌کنم.

فهرست مطالب

عنوان	صفحه
تشکر و قدردانی.....	5
فهرست مطالب.....	7
فهرست شکل‌ها.....	9
چکیده.....	10
فصل اول.....	11
2.1 معرفی محل کارآموزی.....	11
2.1.1 مقدمه.....	11
2.1.2 فعالیت‌ها و ویژگی‌های آزمایشگاه ARAS.....	11
فصل دوم.....	13
3.1 کارهای انجام شده در دوره کارآموزی.....	13
3.1.1 مقدمه.....	13
3.2 معرفی مجموعه داده و نمونه‌ها.....	14
3.2.1 مرحله برش (Incision).....	14
3.2.2 مرحله تزریق ویسکوالاستیک (Viscoelastic Injection).....	14
3.2.3 مرحله کپسولورکس (Capsulorhexis).....	15
3.2.4 مرحله فیکوآمولسیفیکیشن (Phacoemulsification).....	16
3.2.5 مرحله شست‌وشو و مکش (Irrigation/Aspiration).....	16
3.2.6 مرحله کاشت عدسی مصنوعی (Lens Implantation).....	17
3.2.7 مراحل تکمیلی (Lens Positioning, Polishing, Antibiotics).....	18
3.2.8 رفلکس قرمز (Red Reflex) در جراحی آب مروارید.....	18
3.2.9 استخراج فریم‌ها و معیار برچسب‌گذاری بر اساس رفلکس قرمز.....	19
3.3 افزایش داده (Data Augmentation).....	23
3.3.1 آماده‌سازی پایه: پد تا مربع و تغییر اندازه.....	23
3.3.2 تبدیلات هندسی (Geometric).....	24
3.3.3 تبدیلات فوتومتریک (Photometric).....	24
3.3.4 تبدیل به تنسور و نرمال‌سازی.....	24
3.3.5 خط‌مشی طراحی و جرایبی انتخاب پارامترها.....	25
3.4 مدل‌های پیشنهادی برای آموزش و نتایج آن.....	26
3.4.1 شبکه ResNet18.....	26

28.....	ResNet50	3.4.2
29.....	EfficientNet-B0	3.4.3
30.....	EfficientNet-B1	3.4.4
32.....	EfficientNet-B2	3.4.5
33.....	EfficientNet-B3	3.4.6
34.....	EfficientNet-B4	3.4.7
35.....	DenseNet121	3.4.8
37.....	DenseNet169	3.4.9
38.....	VGG16	3.4.10
40.....	VGG19	3.4.11
41.....	جمع‌بندی	3.4.12
43	مراجع	

فهرست شکل‌ها

صفحه	عنوان
14	شکل 1 Incision
15	شکل 2 Viscoelastic
15	شکل 3 Capsulorhexis
16	شکل 4 Phacoemulsification
17	شکل 5 Irrigation Aspiration
17	شکل 6 Lens Implantation
19	شکل 7 Low Nuclear Density
20	شکل 8 Moderate Nuclear Density
20	شکل 9 High Nuclear Density
21	شکل 10 Mature
22	شکل 11 Poor Dilatation
23	شکل 12 Class Distribution
25	شکل 13 Sample Augmentation
27	شکل 14 RESNET18 Loss per Epoch
28	شکل 15 RESNET50 Loss per Epoch
30	شکل 16 EFFICIENTNET-B0 Loss per Epoch
31	شکل 17 EFFICIENTNET-B1 Loss per Epoch
32	شکل 18 EFFICIENTNET-B2 Loss per Epoch
34	شکل 19 EFFICIENTNET-B3 Loss per Epoch
35	شکل 20 EFFICIENTNET-B4 Loss per Epoch
36	شکل 21 DENSENET121 Loss per Epoch
38	شکل 22 DENSENET169 Loss per Epoch
39	شکل 23 VGG16 Loss per Epoch
40	شکل 24 VGG19 Loss per Epoch

چکیده

در این کارآموزی که در آزمایشگاه رباتیک ارس دانشگاه خواجه نصیرالدین طوسی انجام شد، تمرکز بر توسعه یک سامانه هوش مصنوعی برای تعیین سختی جراحی آب مروارید بود. در این پروژه ابتدا مجموعه داده‌ای شامل ۲۰۰۰ تصویر از پنج کلاس «کم»، «متوسط»، «شدید»، «آب مروارید رسیده»^۱ و «تساع ناکافی مردمک»^۲ گردآوری و برچسب‌گذاری گردید. سپس مراحل پیش‌پردازش و آماده‌سازی داده‌ها انجام شد و برای بهبود عملکرد مدل، از روش‌های مختلف افزایش داده استفاده شد. در ادامه، چندین شبکه عصبی عمیق شامل دو مدل ResNet، پنج مدل EfficientNet، دو مدل DenseNet و دو مدل VGG16 و VGG19 با تنظیمات و پیش‌پردازش‌های گوناگون آموزش داده شدند. نتایج به‌دست‌آمده مورد تحلیل قرار گرفت و در نهایت دو مدل با عملکرد برتر به‌عنوان مدل‌های پیشنهادی انتخاب شدند.

واژه‌های کلیدی: هوش مصنوعی، شبکه‌های کانوولوشنی^۳، یادگیری عمیق، شبکه عصبی، جراحی آب مروارید.

¹ Mature Cataract

² Poor Dilation

³ Concolutional Neural Network

1 فصل اول

1.1 معرفی محل کارآموزی

1.1.1 مقدمه

آزمایشگاه سامانه‌های رباتیک پیشرفته و خودکار (ARAS) در دانشکده مهندسی برق دانشگاه صنعتی خواجه نصیرالدین طوسی از سال ۱۳۷۶ فعالیت خود را آغاز کرده است و امروز به‌عنوان یکی از مراکز برجسته‌ی پژوهشی در زمینه رباتیک و سیستم‌های خودکار در کشور شناخته می‌شود. این آزمایشگاه با بیش از دو دهه تجربه، بستری پویا برای پژوهش و آموزش در حوزه‌های مختلف رباتیک فراهم کرده و نقش مهمی در توسعه فناوری‌های نوین و کاربردهای صنعتی و پزشکی داشته است.

مدیریت این مجموعه بر عهده‌ی دکتر حمید تقی‌راد، استاد تمام دانشگاه و پژوهشگر شناخته‌شده در حوزه رباتیک موازی، سیستم‌های کابلی و کنترل پیشرفته است. ایشان با انتشار صدها مقاله علمی و هدایت پروژه‌های متعدد صنعتی و تحقیقاتی، نقش کلیدی در رشد و شکوفایی این آزمایشگاه ایفا کرده‌اند.

1.1.2 فعالیت‌ها و ویژگی‌های آزمایشگاه ARAS

در این بخش، مشخصات و فعالیت‌های کلیدی آزمایشگاه ARAS به‌عنوان محل کارآموزی معرفی می‌شود:

- تاریخچه و اهداف: تأسیس در سال ۱۳۷۶ با هدف انجام پژوهش‌های پیشرفته در زمینه رباتیک و اتوماسیون و گسترش تعامل دانشگاه با صنعت.
- حوزه‌های تحقیقاتی: فعالیت‌های اصلی آزمایشگاه شامل رباتیک جراحی، رباتیک مستقل، واقعیت ترکیبی در جراحی، ربات‌های کابلی و موازی، و هوش مصنوعی و علم داده است.
- ساختار و گروه‌ها: آزمایشگاه دارای چندین گروه پژوهشی است که هرکدام بر یکی از شاخه‌های تخصصی رباتیک و هوش مصنوعی تمرکز دارند.

- فعالیت‌های آموزشی و علمی: برگزاری دوره‌ها، سمینارها و کارگاه‌های تخصصی در زمینه رباتیک و کنترل، همچنین انتشار مقالات در مجلات و کنفرانس‌های معتبر بین‌المللی.
- پروژه‌های صنعتی و کاربردی: اجرای پروژه‌های متعدد در همکاری با صنایع مختلف کشور، از جمله پروژه‌های مربوط به طراحی و کنترل ربات‌های صنعتی و پزشکی.
- امکانات و تجهیزات: بهره‌مندی از سیستم‌های محاسباتی، سخت‌افزارهای رباتیکی، تجهیزات آزمایشگاهی پیشرفته و بسترهای نرم‌افزاری برای شبیه‌سازی و پیاده‌سازی الگوریتم‌های کنترلی.
- پژوهش و توسعه: تمرکز ویژه بر توسعه فناوری‌های نوین در رباتیک و به‌کارگیری روش‌های هوش مصنوعی در حل مسائل پیچیده مهندسی.

2 فصل دوم

2.1 کارهای انجام شده در دوره کارآموزی

2.1.1 مقدمه

این فصل به تشریح دقیق فعالیت‌های انجام شده در طول دوره کارآموزی اختصاص دارد. در این بخش ابتدا مجموعه داده‌ای که مبنای اصلی کار بود معرفی می‌شود و نمونه‌ها و کلاس‌های مختلف آن بررسی خواهند شد. سپس روند برچسب‌گذاری داده‌ها توضیح داده می‌شود، به‌ویژه اینکه چگونه ابتدا یک مجموعه کوچک‌تر و ساده‌تر مورد استفاده قرار گرفت و مدل‌های اولیه روی آن نتایج مطلوبی به دست آوردند، اما با گسترش حجم و پیچیدگی داده‌ها نیاز به مدل‌های قدرتمندتر احساس شد.

در ادامه به موضوع افزایش داده (Data Augmentation) پرداخته می‌شود که نقش کلیدی در جلوگیری از بیش‌برازش^۴ ایفا کرد و به‌طور مبسوط اهمیت آن و انواع روش‌های به‌کاررفته توضیح داده خواهد شد. پس از آن مدل‌های مختلف شبکه‌های عصبی کانولوشنی از جمله خانواده ResNet، EfficientNet، DenseNet و VGG معرفی شده و نتایج حاصل از هر کدام ارائه می‌شود.

در بخش پایانی این فصل، تمرکز ویژه‌ای بر عملکرد مدل‌های EfficientNet-B2 و EfficientNet-B3 خواهد بود که بهترین نتایج را به دست آوردند، و دلایل انتخاب نسخه B2 به‌عنوان مدل نهایی شرح داده خواهد شد. در انتها نیز جمع‌بندی‌ای از تمامی فعالیت‌ها و دستاوردهای این مرحله ارائه می‌گردد.

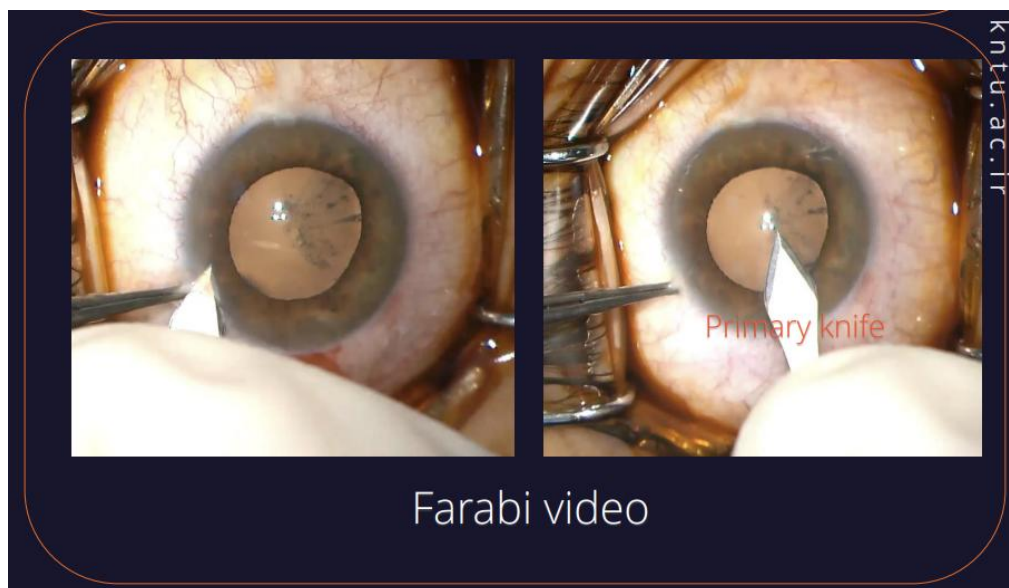
2.2 معرفی مجموعه داده و نمونه‌ها

یکی از مهم‌ترین بخش‌های این پروژه گردآوری مجموعه داده‌ای معتبر و کافی برای آموزش و ارزیابی مدل‌های یادگیری عمیق بود. برای این منظور، حدود ۱۰۰۰ ویدئوی کامل جراحی آب مروارید مورد بازبینی قرار گرفت که در نهایت ۴۳۰ ویدئو با کیفیت مناسب و پوشش کامل مراحل مختلف جراحی انتخاب شدند. از این ویدئوها نمونه‌های تصویری استخراج گردید که در ادامه به عنوان داده‌های ورودی مورد استفاده قرار گرفتند.

جراحی آب مروارید فرایندی چندمرحله‌ای است که هر فاز آن نقش مشخصی در پیشبرد عمل دارد. در ادامه مراحل اصلی این جراحی معرفی شده‌اند.

2.2.1 مرحله برش (Incision)

در آغاز عمل، جراح برش‌های کوچکی بر سطح قرنیه یا صلبیه ایجاد می‌کند. این برش‌ها امکان ورود ابزارهای جراحی به درون چشم را فراهم می‌سازند. طراحی برش‌ها باید به گونه‌ای باشد که هم فضای کافی برای ابزارها ایجاد کند و هم پایداری ساختاری چشم حفظ شود.

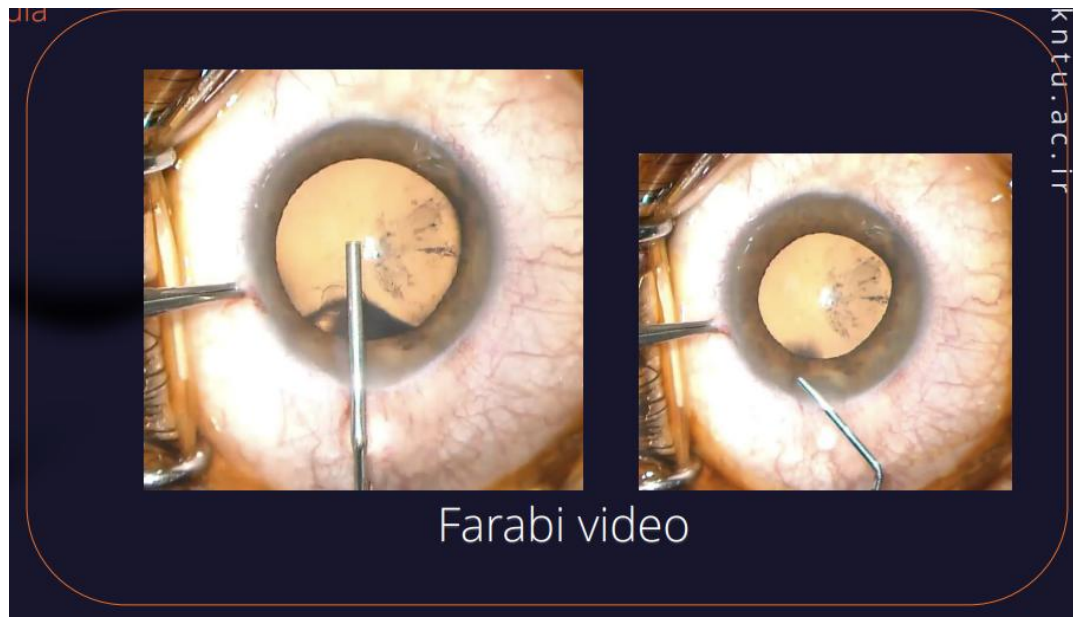


شکل 1 Incision

2.2.2 مرحله تزریق ویسکوالاستیک (Viscoelastic Injection)

در این مرحله، ماده‌ای شفاف و ژل مانند به نام ویسکوالاستیک به درون اتاق قدامی چشم تزریق می‌شود. هدف از این کار ایجاد فضا و محافظت از بافت‌های حساس مانند قرنیه و عدسی است. ویسکوالاستیک علاوه بر ایجاد فاصله کافی برای مانور ابزارهای جراحی، از آسیب مکانیکی به سلول‌های قرنیه جلوگیری کرده و محیطی پایدار

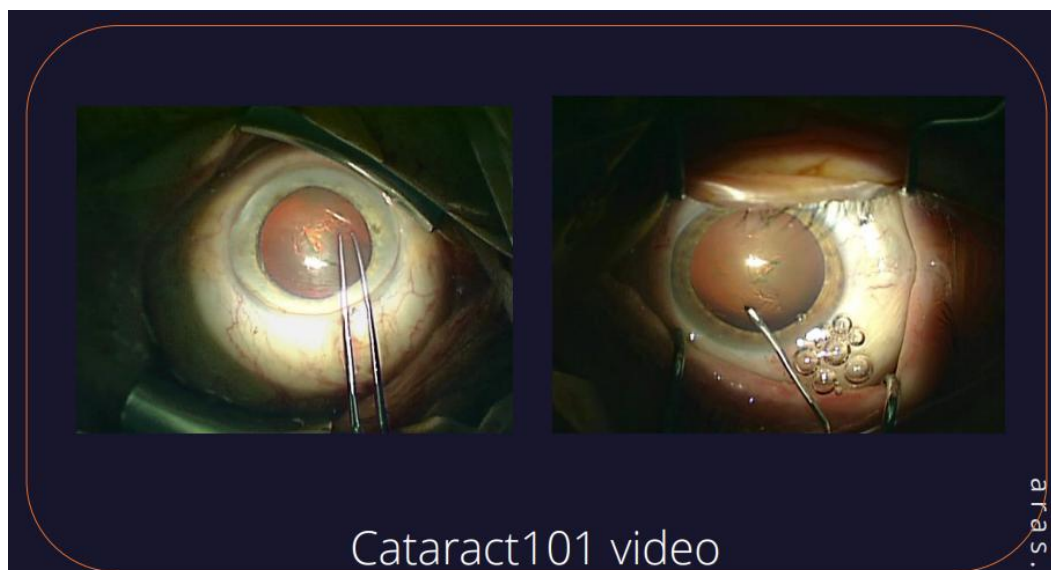
برای انجام مراحل بعدی عمل فراهم می‌کند.



شکل 2 Viscoelastic

2.2.3 مرحله کپسولورکسس (Capsulorhexis)

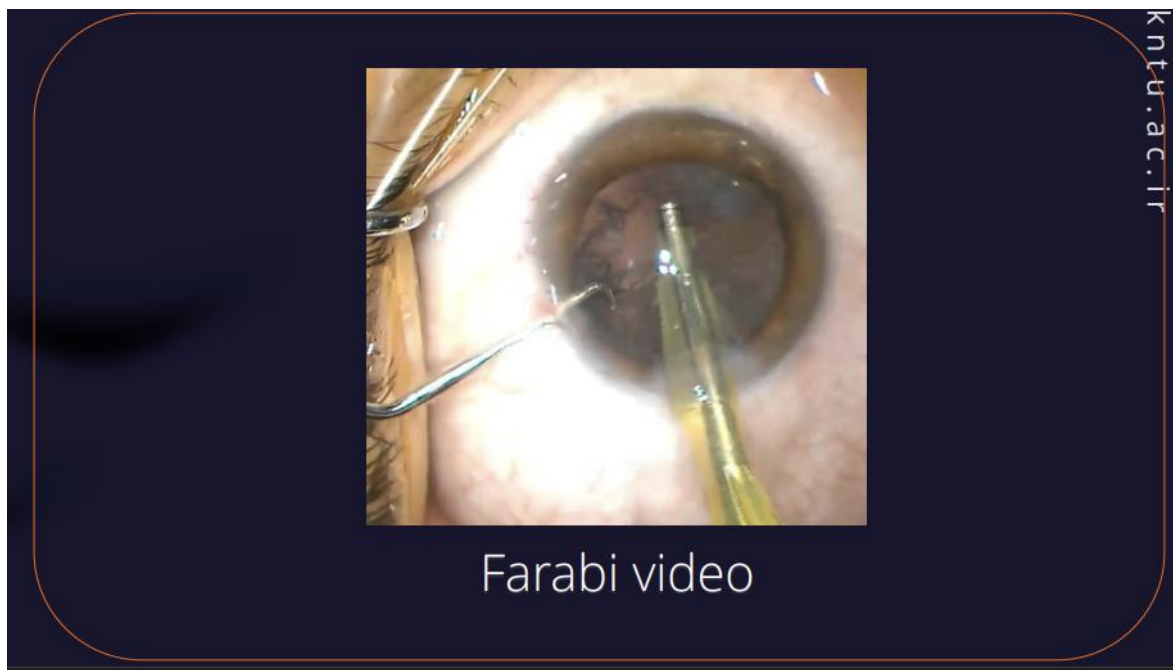
کپسولورکسس یکی از حساس‌ترین بخش‌های جراحی است. در این مرحله، جراح یک برش دایره‌ای منظم در کپسول قدامی عدسی ایجاد می‌کند. این کار باعث می‌شود دسترسی مستقیم به هسته عدسی فراهم شود. کیفیت و یکنواختی این برش نقش تعیین‌کننده‌ای در موفقیت مراحل بعدی عمل و جایگذاری عدسی مصنوعی دارد.



شکل 3 Capsulorhexis

2.2.4 مرحله فیکوآمولسیفیکیشن (Phacoemulsification)

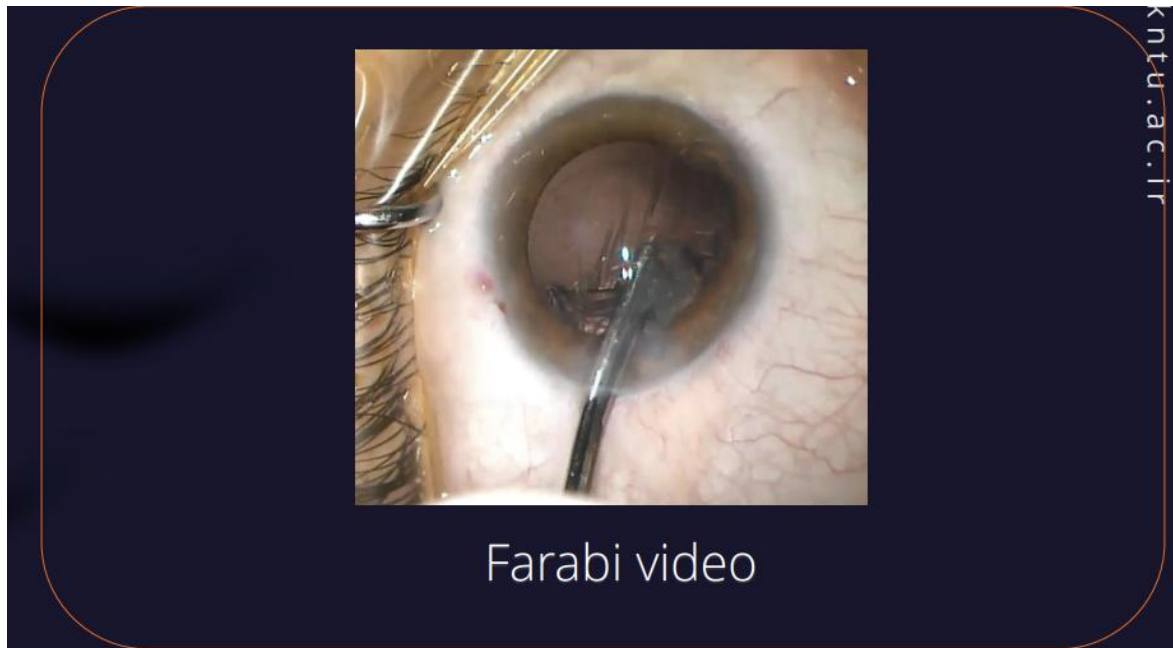
در این مرحله، با استفاده از یک پروب اولتراسوند، هسته سخت عدسی به ذرات ریزتر خرد و به صورت امولسیون در می‌آید. سپس این ذرات با سیستم مکش خارج می‌شوند. فیکوآمولسیفیکیشن مهم‌ترین مرحله در حذف عدسی کدر شده است و نیازمند دقت بالا برای جلوگیری از آسیب به کپسول خلفی و سایر ساختارهای ظریف چشم است.



شکل 4 Phacoemulsification

2.2.5 مرحله شست‌وشو و مکش (Irrigation/Aspiration)

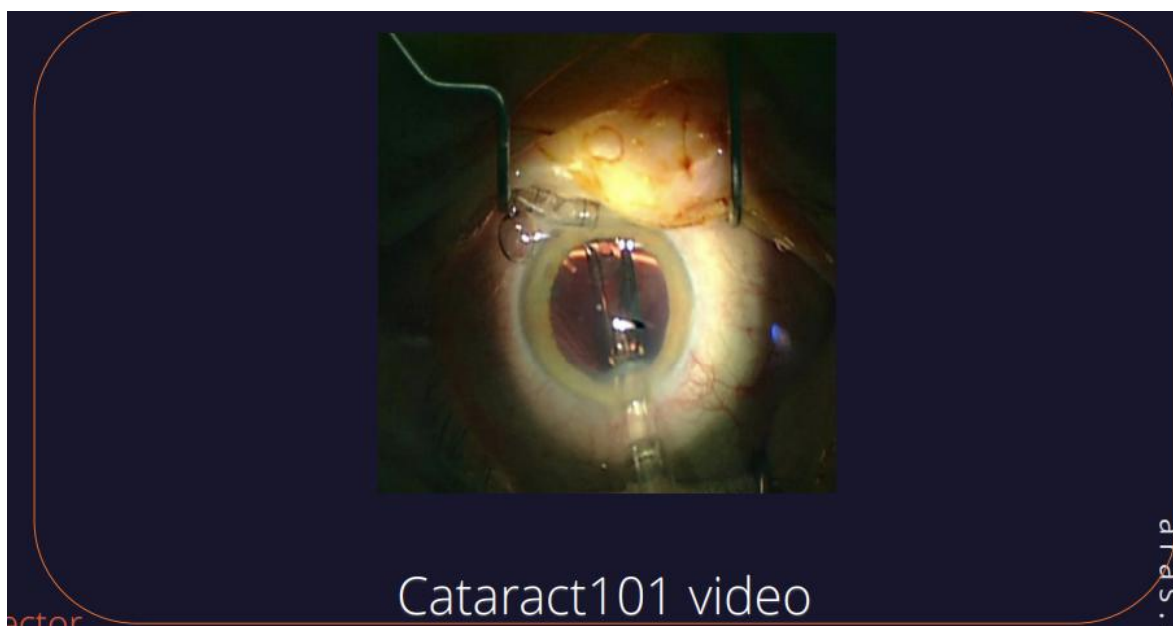
پس از حذف بخش عمده عدسی، بقایای مواد قشری و ذرات باقیمانده باید پاکسازی شوند. این کار با استفاده از ابزارهای شست‌وشو (Irrigation) و مکش (Aspiration) انجام می‌شود. این مرحله برای شفافیت بینایی آینده بیمار حیاتی است، زیرا باقی‌ماندن هرگونه ماده کدر می‌تواند سبب کاهش کیفیت بینایی یا بروز عوارض پس از عمل شود.



شکل 5 Irrigation Aspiration

2.2.6 مرحله کاشت عدسی مصنوعی (Lens Implantation)

پس از پاکسازی کامل، یک عدسی مصنوعی داخل چشمی (IOL) درون کپسول خالی قرار داده می‌شود. این عدسی جایگزین عدسی طبیعی برداشته‌شده می‌گردد و وظیفه تمرکز نور بر شبکیه را بر عهده دارد. انتخاب صحیح نوع و اندازه عدسی تأثیر زیادی بر کیفیت بینایی پس از عمل دارد.



شکل 6 Lens Implantation

2.2.7 مراحل تکمیلی (Lens Positioning, Polishing, Antibiotics)

در پایان عمل، عدسی کاشته شده در جای مناسب تثبیت می شود. سپس کپسول و محیط داخلی چشم پولیش یا پاک سازی نهایی می گردد تا از باقی ماندن ذرات یا رسوبات جلوگیری شود. در نهایت، داروهای آنتی بیوتیک برای پیشگیری از عفونت تزریق می شوند. این اقدامات تکمیلی تضمین کننده موفقیت بلندمدت عمل هستند.

2.2.8 رفلکس قرمز (Red Reflex) در جراحی آب مروارید

یکی از پدیده های مهمی که در طول جراحی آب مروارید توسط دوربین یا میکروسکوپ جراحی قابل مشاهده است، رفلکس قرمز می باشد. این پدیده زمانی رخ می دهد که نور تابیده شده از میکروسکوپ از طریق قرنیه و عدسی به سمت شبکیه عبور کرده و سپس از شبکیه بازتاب پیدا می کند. به دلیل غنی بودن شبکیه از عروق خونی، این بازتاب به رنگ قرمز مشاهده می شود.

رفلکس قرمز در مراحل مختلف جراحی دیده می شود، اما بیشترین اهمیت آن در ابتدای مرحله کپسولورکسس (Capsulorhexis) است. در این مرحله، جراح نیاز دارد تا مرزها و شفافیت عدسی را با دقت مشاهده کند تا بتواند یک برش دایره ای دقیق و یکنواخت در کپسول قدامی ایجاد نماید. مشاهده ی شدت و کیفیت رفلکس قرمز به جراح کمک می کند تا درک مناسبی از میزان کدورت عدسی و مرحله پیشرفت آب مروارید داشته باشد.

به بیان دیگر، در بیماران با آب مروارید خفیف یا متوسط، رفلکس قرمز روشن و واضح است؛ اما هرچه شدت و سختی کاتاراکت بیشتر شود (به ویژه در مراحل رسیده یا متراکم)، این رفلکس کاهش می یابد یا حتی محو می شود. بنابراین، ارزیابی رفلکس قرمز در ابتدای کپسولورکسس یکی از ابزارهای غیرمستقیم برای سنجش سطح پیشرفت آب مروارید محسوب می شود.

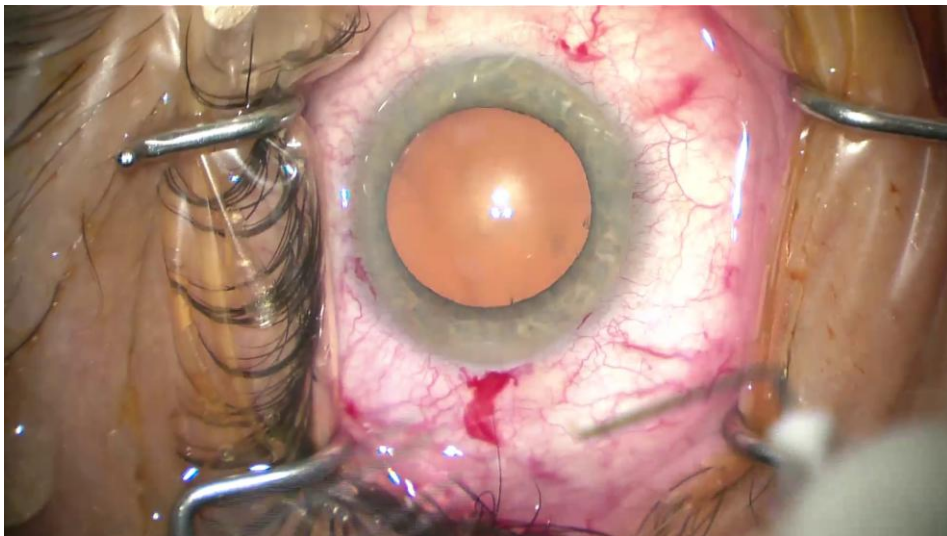
از دیدگاه پردازش تصویر نیز، رفلکس قرمز می تواند به عنوان یک ویژگی کلیدی در طبقه بندی مراحل مختلف کاتاراکت مورد توجه قرار گیرد، زیرا شدت آن رابطه مستقیمی با میزان کدورت عدسی و سختی جراحی دارد.

2.2.9 استخراج فریم‌ها و معیار برچسب‌گذاری بر اساس رفلکس قرمز

برای ارزیابی مرحله‌ی کاتاراکت و تشکیل کلاس‌های مختلف در مجموعه داده، از بخش کیسولورکسس به عنوان مهم‌ترین مرحله جراحی استفاده شد. از این مرحله، فریم‌ها با نرخ ۵ فریم بر ثانیه استخراج گردیدند. سپس به صورت دستی فریم‌هایی که در آن‌ها رفلکس قرمز (Red Reflex) به وضوح قابل مشاهده بود (در صورت وجود) انتخاب شدند.

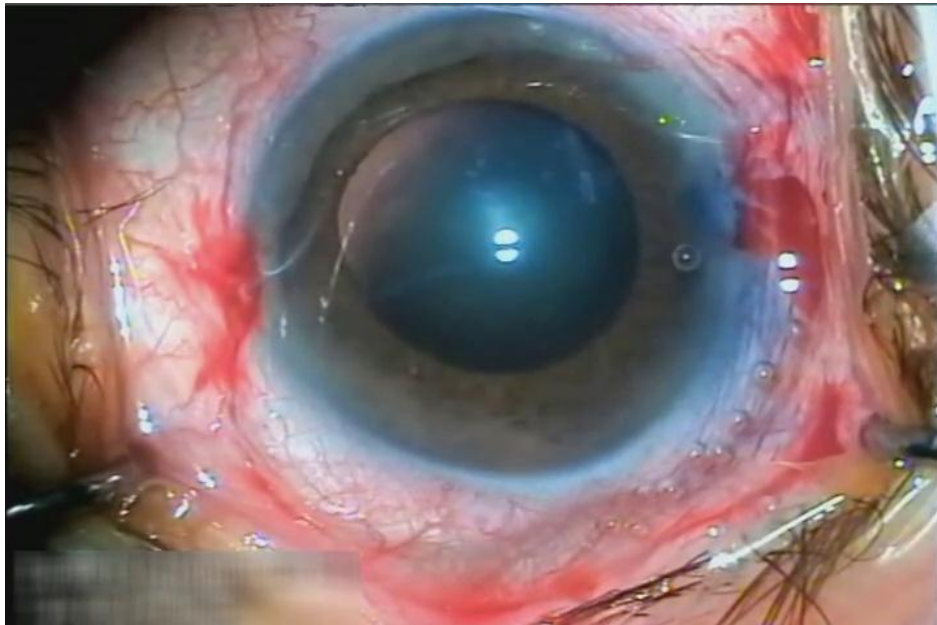
مبنای اصلی برچسب‌گذاری، میزان و شدت رفلکس قرمز عبوری از عدسی بود. بر اساس کیفیت مشاهده شده، سه کلاس اصلی «کم»، «متوسط» و «شدید» تعیین شدند:

- کلاس کم (Low Nuclear Density): در این حالت، رفلکس قرمز در تمام سطح عدسی چشم قابل مشاهده است. عبور نور از عدسی نشان‌دهنده‌ی شفافیت نسبی و کدورت خفیف هسته است.



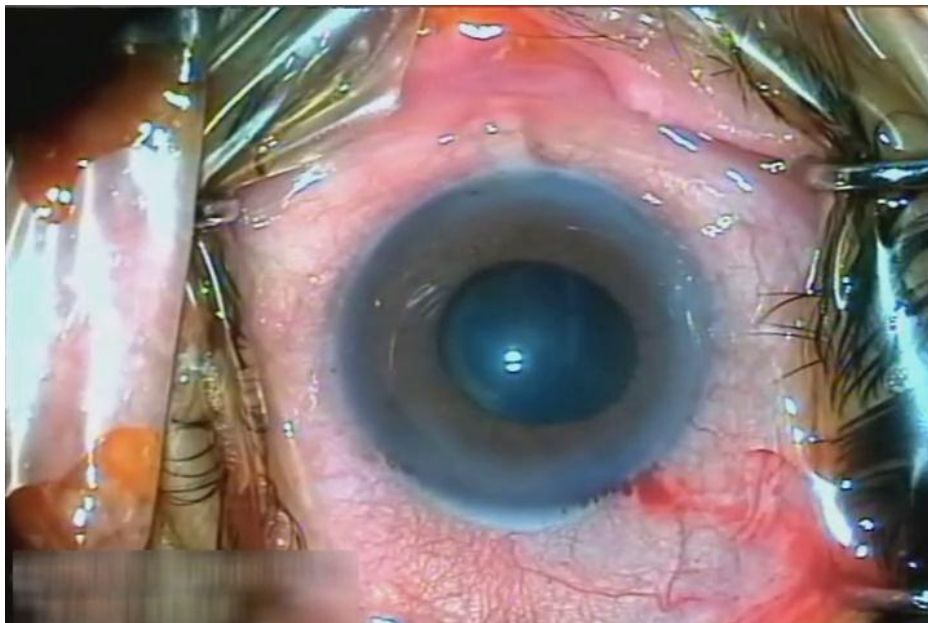
شکل 7 Low Nuclear Density

- کلاس متوسط (Moderate Nuclear Density): در این حالت تنها بخش‌هایی از عدسی اجازه عبور رفلکس قرمز را می‌دهند. کدورت هسته بیشتر از حالت قبل است و نواحی تاریک‌تر در کنار نواحی شفاف دیده می‌شوند



شکل 8 Moderate Nuclear Density

- کلاس شدید (High Nuclear Density): در این حالت هیچ رفلکس قرمزی از عدسی عبور نمی‌کند. هسته‌ی عدسی به‌طور کامل کدر است و این موضوع بیانگر پیشرفت شدید کاتاراکت است.



شکل 9 High Nuclear Density

• کلاس چهارم: آب مروارید رسیده (Mature Cataract)

علاوه بر سه دسته‌ی فوق، کلاس دیگری تحت عنوان آب مروارید رسیده تعریف شد. در این حالت، هسته عدسی به شدت متراکم شده و لایه‌ی قشری (Cortex) حالت شیری و کدر پیدا می‌کند. چشم در این شرایط به صورت سفید یا متمایل به خاکستری دیده می‌شود و حتی در صورت تابش نور، هیچ رفلکس قرمزی قابل مشاهده نیست. این وضعیت نشان‌دهنده‌ی یک آب مروارید در مراحل نهایی است که عمل جراحی را به مراتب دشوارتر می‌سازد.



شکل 10 Mature

• کلاس پنجم: اتساع ناکافی مردمک (Poor Dilation)

آخرین کلاس تعریف‌شده مربوط به مواردی است که مردمک چشم به اندازه‌ی کافی باز نشده باشد. در این حالت، جراح با محدودیت دید و دسترسی روبه‌رو می‌شود و حتی در صورت شفافیت نسبی عدسی، رفلکس قرمز یا سایر ویژگی‌ها به خوبی دیده نمی‌شوند.

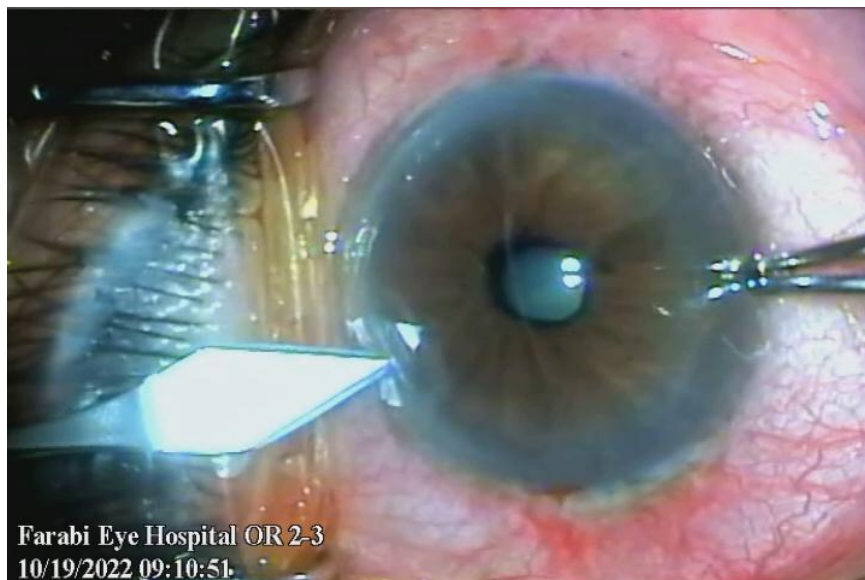
دلایل اتساع ناکافی مردمک می‌تواند شامل موارد زیر باشد:

- تغییرات وابسته به سن که منجر به کاهش پاسخ مردمک به داروهای گشادکننده می‌شود.
- مصرف داروهای خاص مانند آلفابلاکرها.
- وجود بیماری‌های چشمی زمینه‌ای مثل سندرم شبه‌اکسفولیشن

(Pseudoexfoliation Syndrome)

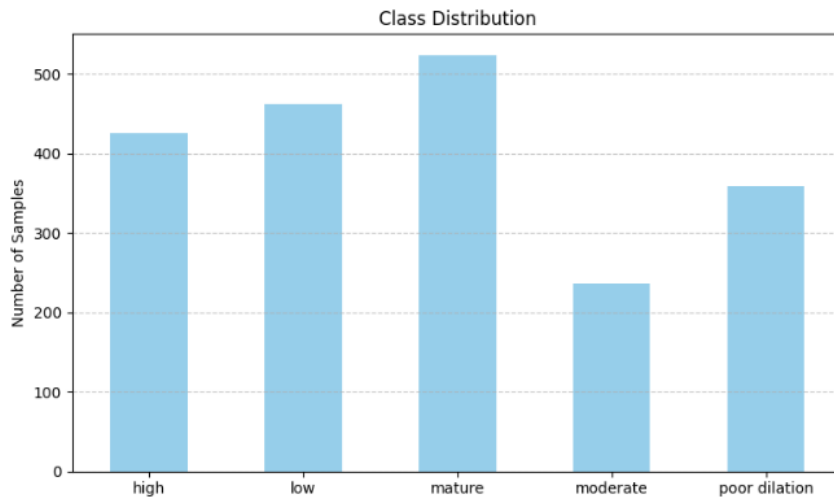
- وجود بافت‌های فیبروتیک یا چسبندگی در ناحیه عنبیه.

این وضعیت نیز به‌عنوان یک کلاس مستقل در نظر گرفته شد، زیرا به‌شدت بر دشواری عمل جراحی تأثیر می‌گذارد



شکل 11 Poor Dilation

- در پایان، پس از طی تمامی مراحل شامل انتخاب ویدئوها، استخراج فریم‌ها، مشاهده و ارزیابی رفلکس قرمز، و برچسب‌گذاری دستی تصاویر بر اساس معیارهای تعریف‌شده، مجموعه داده نهایی آماده گردید. نتیجه این فرایند، تشکیل یک پایگاه داده شامل پنج کلاس اصلی بود که توزیع آن در میان دسته‌های مختلف به‌صورت شکل صفحه بعد است:



شکل 12 Class Distribution

2.3 افزایش داده (Data Augmentation)

هدف از افزایش داده، افزایش تنوع آماری نمونه‌ها بدون تغییر معنای بالینی تصویر است تا مدل در برابر تغییرات رایج (چرخش‌های جزئی میکروسکوپ، برش و جابجایی خفیف میدان دید، تغییرات نور و کنتراست، و اندکی تاری حرکت) مقاوم شود و از بیش‌برازش روی الگوهای ظاهری ثابت جلوگیری گردد. در همه‌ی مواردی که نیاز به «رنگ پس‌زمینه» برای پُر کردن نواحی خالی بوده، به‌جای مشکلی غیرواقعی از میانگین رنگ تصویر استفاده شده تا آرتیفکت مرزی ایجاد نشود و ویژگی‌های بالینی (مانند رفلکس قرمز، لبه‌ی کپسولورکسس، بافت قشر عدسی) مخدوش نگردد. افزونه‌ها فقط روی مجموعه‌ی آموزش اعمال شده و برای اعتبارسنجی/آزمون، صرفاً تبدیل‌های تعیینی (پد، تغییر اندازه، نرمال‌سازی) به‌کار رفته است.

2.3.1 آماده‌سازی پایه: پد تا مربع و تغییر اندازه

- **PadToSquareWithMean**: تصویر را با رنگ میانگین خودش تا مربع پد می‌کند تا نسبت تصویر یکنواخت شود و از کشیدگی‌های مخرب جلوگیری گردد.

```
class PadToSquareWithMean:
```

```
    def __call__(self, image):
```

```
        w, h = image.size; m = max(w, h)
```

```
        pl = (m - w) // 2; pt = (m - h) // 2
```

```
        pr = m - w - pl; pb = m - h - pt
```

```
        mean_color = _mean_rgb(image)
```

تصویر RGB میانگین

```
return TF.pad(image, (pl, pt, pr, pb), fill=mean_color)
```

- `Resize(260×260)`: همه‌ی ورودی‌ها به اندازه‌ی ثابت 260×260 پیکسل تغییر اندازه داده می‌شوند تا با ورودی شبکه‌های از پیش‌آموزه سازگار شوند.

```
transforms.Resize((260, 260))
```

2.3.2 تبدیلات هندسی (Geometric)

- `RotateWithMeanFill ($\pm 12^\circ$)`: چرخش‌های خفیف میکروسکوپ یا بیمار را شبیه‌سازی می‌کند؛ پرشدن با میانگین رنگ مانع ایجاد نوارهای سیاه می‌شود.

```
class RotateWithMeanFill:
```

```
def __call__(self, image):
```

```
    angle = transforms.RandomRotation.get_params([-12, 12])
```

```
    return TF.rotate(image, angle, fill=_mean_rgb(image))
```

2.3.3 تبدیلات فوتومتریک (Photometric)

- `ColorJitter (p=0.6)` با `brightness=0.25`, `contrast=0.25`, `saturation=0.10`: تغییرات نور محیط/اکسپوژر دوربین جراحی را مدل می‌کند، اما دامنه طوری انتخاب شده که نشانه‌های بالینی مانند شدت رفلکس قرمز تحریف نشوند.

```
transforms.RandomApply([
```

```
    transforms.ColorJitter(brightness=0.25, contrast=0.25, saturation=0.10)
```

```
], p=0.6)
```

2.3.4 تبدیل به تنسور و نرمال‌سازی

- `ToTensor`: تبدیل تصویر به قالب CHW و بازه‌ی $[0,1]$.

```
transforms.ToTensor()
```

```
IMAGENET_MEAN = [0.485, 0.456, 0.406]
```

```
IMAGENET_STD = [0.229, 0.224, 0.225]
```

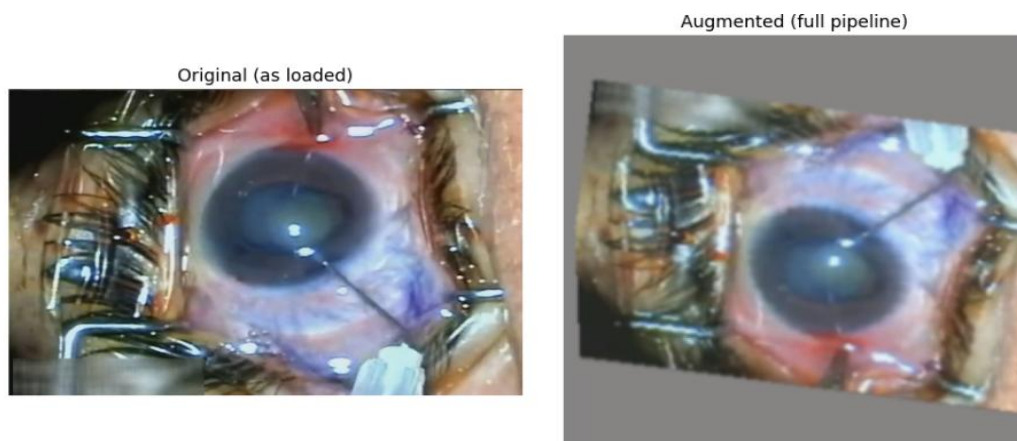
```
transforms.Normalize(mean=IMAGENET_MEAN, std=IMAGENET_STD)
```


2.3.5 خطمشی طراحی و چرایی انتخاب پارامترها

- دامنه‌های چرخش/جابجایی/برش عمداً محدود انتخاب شدند تا شکل آناتومیک مردمک، لبه‌ی برش کپسول و الگوی رفلکس قرمز دچار دفرم جدی نشود.
- احتمال‌ها (p) طوری تنظیم شد که در هر ایپاک، ترکیبی واقع‌گرایانه اما متنوع از نمونه‌ها دیده شود؛ افزایش بی‌رویه‌ی فوتومتریک‌ها می‌تواند به «یادگیری میان‌بر» و کاهش حساسیت مدل به سیگنال‌های بالینی منجر شود.
- پر کردن نواحی خارج از قاب با میانگین رنگ تصویر، نسبت به رنگ مشکی، از ایجاد مرزهای غیرطبیعی جلوگیری می‌کند و به‌ویژه هنگام تشخیص شدت رفلکس قرمز، نرخ خطا را کاهش می‌دهد.
- بر مجموعه‌ی اعتبارسنجی/آزمون، هیچ‌یک از تبدیلات تصادفی اعمال نشد تا ارزیابی بی‌طرفانه باقی بماند؛ تنها $\text{Pad} \rightarrow \text{Resize} \rightarrow \text{ToTensor} \rightarrow \text{Normalize}$ استفاده شد.

برای درک بهتر تأثیر افزایش داده، در شکل زیر یک نمونه از تصویر خام و نسخه‌ی افزایش‌یافته‌ی آن ارائه شده است. در تصویر سمت چپ نمونه‌ی اصلی بدون هیچ تغییری نمایش داده شده و در تصویر سمت راست همان تصویر پس از عبور از کل زنجیره‌ی افزایش داده دیده می‌شود.

در نسخه‌ی افزایش‌یافته می‌توان مشاهده کرد که تغییراتی همچون چرخش خفیف، جابجایی میدان دید، و تنظیم رنگ و روشنایی اعمال شده‌اند. علاوه بر این، پس‌زمینه‌ی تصویر با میانگین رنگ ناحیه پر شده تا از ایجاد مرزهای غیرواقعی جلوگیری گردد. این نوع تغییرات نه‌تنها ماهیت بالینی تصویر را حفظ می‌کنند (مانند رفلکس قرمز و مرزهای مردمک)، بلکه باعث افزایش تنوع نمونه‌ها می‌شوند و در نتیجه مدل در برابر تغییرات ناخواسته در شرایط واقعی جراحی مقاوم‌تر خواهد شد.



شکل 13 Sample Augmentation

2.4 مدل های پیشنهادی برای آموزش و نتایج آن

پس از آماده سازی و برچسب گذاری مجموعه داده، مرحله ی بعدی، انتخاب و آموزش مدل های یادگیری عمیق برای طبقه بندی تصاویر استخراج شده از مراحل جراحی آب مروارید بود. در این پروژه، از چندین معماری مطرح در حوزه ی بینایی ماشین استفاده شد تا عملکرد آن ها در مواجهه با داده های پزشکی بررسی و مقایسه شود. هدف اصلی این بخش، ارزیابی توانایی مدل ها در استخراج الگوهای تصویری مرتبط با شدت کاتاراکت و شرایط خاص چشم، و همچنین مقایسه ی دقت، تعادل میان کلاس ها، و کارایی محاسباتی آن ها بود. برای این منظور، مجموعه ای از شبکه های کانولوشنی شامل خانواده ی ResNet و EfficientNet انتخاب شدند. هر مدل ابتدا معرفی و معماری آن توضیح داده می شود، سپس روند آموزش و تنظیمات مربوطه شرح داده خواهد شد، و در نهایت نتایج عددی و تحلیلی به تفکیک ارائه می گردد.

2.4.1 شبکه ResNet18

شبکه ی RESNET18 یکی از اعضای خانواده ی ResNet است که نخستین بار در سال ۲۰۱۵ معرفی شد و نقطه ی عطفی در یادگیری عمیق برای بینایی ماشین به شمار می رود. ایده ی اصلی این معماری استفاده از ارتباطات باقیمانده (Residual Connections) یا همان *skip connections* است. در شبکه های کانولوشنی سنتی، هرچه عمق شبکه بیشتر می شود، مشکلاتی مانند محو شدن گرادیان و کاهش دقت آموزشی رخ می دهد. ResNet این مشکل را با افزودن مسیرهای میان بر بین لایه ها حل می کند تا شبکه بتواند نگاشت همانی (Identity Mapping) را به راحتی یاد بگیرد. همین ویژگی باعث شد آموزش شبکه های بسیار عمیق ممکن شود.

نسخه ی ResNet18 شامل ۱۸ لایه ی قابل یادگیری است و از بلوک های باقیمانده ی متوالی تشکیل شده است. این معماری به دلیل سادگی و سبکی، در مقایسه با نسخه های عمیق تر (مانند ResNet50 یا RESNET101) سریع تر آموزش داده می شود، ولی همچنان عملکرد مناسبی ارائه می دهد. این مدل به صورت اولیه روی مجموعه داده ی ImageNet آموزش دیده است؛ بنابراین ویژگی های عمومی تصاویر (لبه ها، اشکال و بافت ها) را از پیش آموخته و قابلیت استفاده در یادگیری انتقالی را دارد.

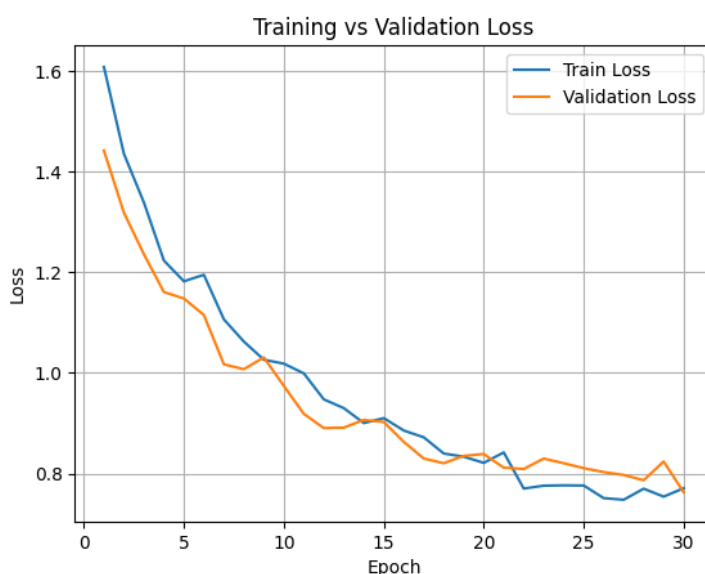
در این پروژه، از نسخه ی از پیش آموزش دیده ی ResNet18 استفاده شد و لایه ی آخر آن که در حالت استاندارد ۱۰۰۰ کلاس ImageNet را پیش بینی می کند، با یک لایه ی تمام متصل جدید جایگزین شد تا پنج کلاس مربوط به مراحل کاتاراکت را پیش بینی کند: کم تراکم (Low)، متوسط (Moderate)، پرتراکم (High)، رسیده (mature) و اتساع ناکافی مردمک (poor dilation). در مراحل اولیه، همه ی لایه ها منجمد شدند تا ویژگی های عمومی استخراج شده از ImageNet حفظ گردد و تنها لایه ی آخر آموزش داده شود. سپس در حالت های خاص، لایه های بالاتر (مانند بلوک چهارم و لایه ی نهایی) برای ریزتنظیم (Fine-Tuning) باز شدند تا شبکه بتواند

ویژگی‌های خاص‌تر داده‌های پزشکی را نیز یاد بگیرد.

برای آموزش، داده‌ها به مجموعه‌های آموزش، اعتبارسنجی و آزمون تقسیم شدند. از توابع افزایش داده (Data Augmentation) برای جلوگیری از بیش‌برازش و افزایش تنوع نمونه‌ها استفاده شد. آموزش شبکه با استفاده از تابع خطای Cross-Entropy، بهینه‌ساز Adam با نرخ یادگیری اولیه ۰.۰۰۰۱ و برنامه‌ریز کاهش نرخ یادگیری (ReduceLROnPlateau) انجام شد. حداکثر ۳۰ اپاک (Epoch) برای آموزش در نظر گرفته شد و از معیار توقف زودهنگام (Early Stopping) با صبر ۵ اپاک استفاده گردید تا زمانی که شبکه پیشرفتی در خطای اعتبارسنجی نداشت، آموزش متوقف شود.

در پایان آموزش، عملکرد ResNet18 بر روی مجموعه‌ی آزمون مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل توانسته است به دقت کلی (Overall Accuracy) برابر با 0.7562 دست پیدا کند. همچنین امتیاز F1 ماکرو 0.7470 و امتیاز F1 وزن‌دار 0.7575 حاصل شد که نشان‌دهنده‌ی تعادل قابل‌قبول در تشخیص کلاس‌ها است. تحلیل دقیق‌تر معیارها نشان داد که مدل در تشخیص کلاس کم‌تراکم $Precision=0.8511$ ، $Recall=0.9091$ و اتساع ناکافی مردمک $Precision=0.8636$ ، $Recall=0.703$ عملکرد مناسبی داشت. در مقابل، عملکرد آن در تفکیک کلاس‌های «پرتراکم» و «رسیده» ضعیف‌تر بود، به‌ویژه در کلاس پرتراکم که دقت (Precision) به حدود 0.6176 کاهش یافت. این موضوع می‌تواند ناشی از شباهت بصری بالا بین مراحل پیشرفته‌ی کاتاراکت باشد.

همانطور که در شکل 14 می‌بینیم، منحنی Loss بر حسب Epoch رسم شده که نشان دهنده یادگیری پیوسته این مدل می‌باشد.



شکل 14 RESNET18 Loss per Epoch

به طور کلی، نتایج ResNet18 نشان داد که این مدل سبک و سریع قادر است الگوهای کلیدی برای طبقه‌بندی تصاویر کاتاراکت را یاد بگیرد و عملکرد قابل قبولی ارائه دهد. هرچند برای بهبود بیشتر، استفاده از معماری‌های پیچیده‌تر مانند EfficientNet پیشنهاد می‌شود.

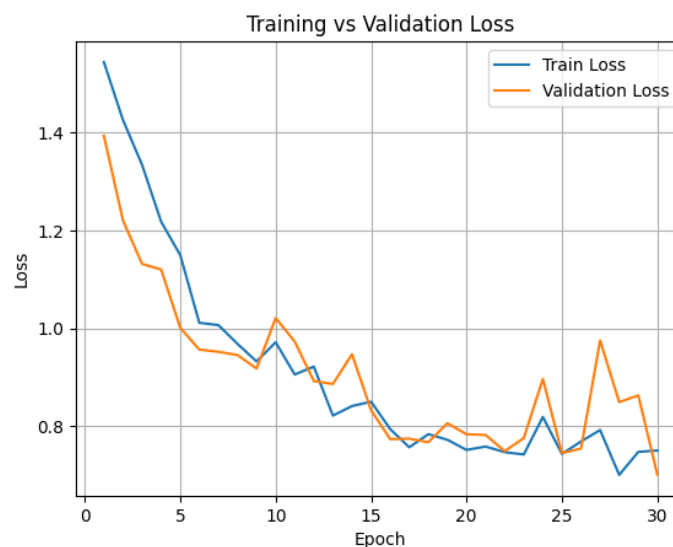
2.4.2 ResNet50

ResNet50 یکی از شناخته‌شده‌ترین معماری‌های شبکه‌های عصبی کانولوشنی عمیق است که بر پایه‌ی ایده‌ی «اتصالات میان‌بری» (Skip Connections) «توسعه یافته و با ۵۰ لایه عمقی‌تر از ResNet18 عمل می‌کند. این معماری به دلیل توانایی بالا در استخراج ویژگی‌های پیچیده‌تر از تصاویر، به‌ویژه در کاربردهای پزشکی، انتخاب مناسبی محسوب می‌شود. استفاده از بلوک‌های Bottleneck در ResNet50 باعث می‌شود که با وجود عمق زیاد، حجم محاسباتی شبکه کنترل شده باقی بماند و مشکل ناپدید شدن گرادپان‌ها کمتر رخ دهد.

در این پروژه، ResNet50 به صورت از پیش آموزش دیده (Pretrained) بر روی دیتاست ImageNet بارگذاری شد و لایه‌ی نهایی آن با یک لایه‌ی خطی جایگزین گردید تا خروجی برای پنج کلاس مورد نظر

(Low, Moderate, High, Mature, Poor Dilation) تولید شود. در طی آموزش، از تکنیک‌هایی همچون تنظیم نرخ یادگیری با استفاده از ReduceLROnPlateau، استفاده از Early Stopping برای جلوگیری از بیش‌برازش، و به کارگیری توزیع متوازن داده‌ها بر اساس وزن‌دهی کلاس‌ها استفاده شد.

فرآیند آموزش طی ۳۰ دوره (Epoch) انجام شد. در نمودار «Training vs Validation Loss» دیده می‌شود که هر دو منحنی با شیب قابل قبول کاهش یافته و در میانه‌ی دوره‌ها به ثبات نسبی رسیدند. این موضوع نشان می‌دهد که مدل توانسته است الگوهای مؤثر را از داده‌ها یاد بگیرد و دچار بیش‌برازش قابل توجه نشده است.



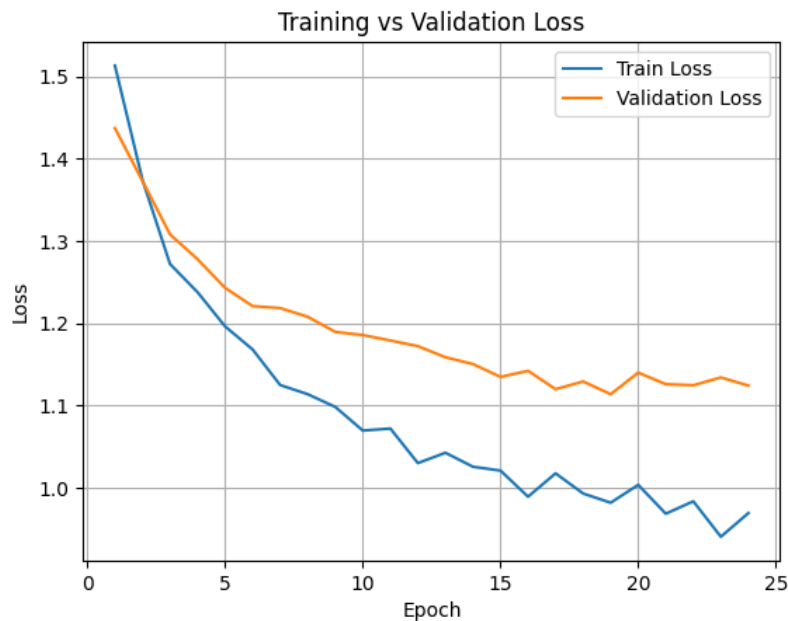
شکل 15 RESNET50 Loss per Epoch

نتایج نهایی ارزیابی ResNet50 به صورت زیر بود: مدل به دقت کلی (Accuracy) برابر با 74.38 درصد دست یافت. میانگین امتیاز Macro F1 برابر با 0.7267 و Weighted F1 برابر با 0.7397 گزارش شد که نشان دهنده تعادل نسبی میان کلاس‌ها است. در سطح کلاس‌ها، بهترین عملکرد مربوط به کلاس Low nuclear density با F1-score برابر با 0.8791 بود، در حالی که کلاس‌های Moderate و Mature دقت پایین‌تری داشتند که عمدتاً ناشی از تعداد نمونه‌های کمتر و شباهت ویژگی‌های بصری با سایر کلاس‌هاست. به طور کلی، ResNet50 در مقایسه با ResNet18 به دلیل عمق بیشتر توانست الگوهای پیچیده‌تری را از داده‌ها استخراج کند، اما افزایش عمق مدل الزاماً به بهبود چشمگیر در تمام کلاس‌ها منجر نشد. این نتیجه بیانگر اهمیت توازن میان عمق شبکه، حجم داده‌ها و میزان متنوع‌سازی از طریق افزایش داده (Data Augmentation) در مسائل پزشکی است.

EfficientNet-B0 2.4.3

EfficientNet-B0 یکی از مدل‌های پایه خانواده EfficientNet است که توسط گوگل معرفی شد و بر اساس ایده‌ی اسکیل متوازن (Scaling) طراحی گردیده است. در این روش، به جای افزایش صرف عمق (Depth) یا عرض (Width) شبکه، مقیاس‌دهی به‌طور همزمان در سه بعد عمق، عرض و رزولوشن ورودی صورت می‌گیرد. همین مسئله موجب شده تا EfficientNet-B0 نسبت به بسیاری از مدل‌های کانولوشنی قدیمی‌تر، کارایی بالاتری در استخراج ویژگی‌ها داشته باشد در حالی که تعداد پارامترها و هزینه محاسباتی آن کمتر باقی می‌ماند. در این پروژه، مدل EfficientNet-B0 به صورت از پیش آموزش‌دیده روی دیتاست ImageNet بارگذاری و لایه‌ی خروجی آن برای پنج کلاس مورد نظر جایگزین گردید. هدف اصلی از استفاده از این معماری، بررسی میزان توانایی آن در شناسایی تفاوت‌های ریزبینانه میان مراحل مختلف کاتاراکت و تشخیص سطوح تراکم عدسی چشم بود. برای آموزش، همانند مدل‌های دیگر، از تنظیم نرخ یادگیری، توقف زودهنگام و وزن‌دهی به کلاس‌ها جهت مقابله با عدم توازن داده‌ها استفاده شد.

نمودار آموزش و اعتبارسنجی نشان می‌دهد که اگرچه خطای آموزش (Train Loss) روند نزولی خوبی داشت، اما خطای اعتبارسنجی (Validation Loss) در سطح بالاتری تثبیت شد و فاصله‌ی نسبی میان این دو منحنی در اواخر دوره‌ها به وجود آمد. این موضوع می‌تواند نشانه‌ای از بروز پدیده بیش‌برازش باشد، به‌ویژه به دلیل ظرفیت محدود مدل B0 در مقایسه با نسخه‌های بزرگ‌تر EfficientNet.



شکل 16 EFFICIENTNET-B0 Loss per Epoch

از نظر نتایج عددی، مدل EfficientNet-B0 به دقت کلی (Accuracy) برابر با 59.52 درصد دست یافت. همچنین امتیاز Macro F1 برابر با 0.6004 و Weighted F1 برابر با 0.5914 گزارش شد. در میان کلاس‌ها، بهترین عملکرد مربوط به کلاس Poor Dilation با F1-score برابر با 0.7660 بود، در حالی که کلاس Moderate به دلیل کمبود داده و شباهت‌های زیاد با سایر کلاس‌ها پایین‌ترین امتیاز (0.5000) را داشت.

به طور کلی، نتایج به دست آمده نشان می‌دهد که هرچند EfficientNet-B0 به عنوان نسخه‌ی پایه این خانواده قادر به استخراج برخی الگوها از داده‌های جراحی چشم است، اما دقت و تعمیم‌پذیری آن در مقایسه با مدل‌های عمیق‌تر یا کارآمدتر محدودتر می‌باشد. این مسئله بیانگر آن است که برای چنین مسئله پزشکی پیچیده‌ای، نسخه‌های بزرگ‌تر مانند B2 یا B3 گزینه‌های بهتری خواهند بود.

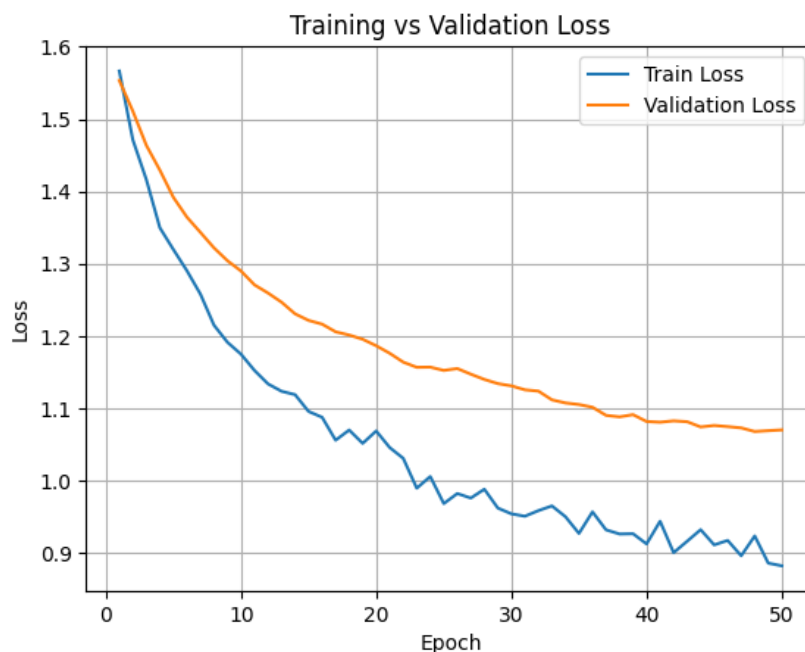
EfficientNet-B1 2.4.4

مدل EfficientNet-B1 نسخه‌ی بزرگ‌تر و بهینه‌تر از B0 است که با افزایش مقیاس‌دهی در ابعاد عمق، عرض و رزولوشن ورودی طراحی شده است. این معماری به دلیل تعداد پارامتر بیشتر و ظرفیت بالاتر، توانایی بهتری در یادگیری ویژگی‌های پیچیده‌تر تصاویر دارد. با این حال، این افزایش ظرفیت نیازمند داده‌های بیشتر و متنوع‌تر برای جلوگیری از بیش‌برازش است.

در این پروژه، EfficientNet-B1 با وزن‌های از پیش آموزش دیده بر روی دیتاست ImageNet بارگذاری شد و لایه‌ی خروجی آن برای پنج کلاس تعریف شده جایگزین گردید. مراحل آموزش مشابه مدل‌های قبلی شامل

استفاده از تکنیک‌هایی مانند کاهش پویا نرخ یادگیری (ReduceLROnPlateau)، توقف زود هنگام و وزن‌دهی به کلاس‌ها بود. هدف اصلی از آزمایش B1 بررسی عملکرد آن در مقایسه با B0 و بررسی تأثیر افزایش ظرفیت مدل بر دقت نهایی بود.

نمودار آموزش و اعتبارسنجی نشان می‌دهد که هر دو منحنی کاهش مناسبی داشتند و اگرچه فاصله میان Train Loss و Validation Loss در دوره‌های پایانی باقی ماند، اما روند کلی نسبتاً پایدار بود. این مسئله نشان می‌دهد که مدل موفق به یادگیری الگوها شده است، هرچند که برای دستیابی به تعمیم‌پذیری بهتر نیاز به داده‌های بیشتر یا افزایش تنوع در افزایش داده‌ها (Augmentation) وجود دارد.



شکل 17 EFFICIENTNET-B1 Loss per Epoch

از نظر عملکرد، مدل EfficientNet-B1 به دقت کلی (Accuracy) برابر با 61.38 درصد دست یافت. همچنین Macro F1 برابر با 0.6109 و Weighted F1 برابر با 0.5982 گزارش شد. بهترین عملکرد مربوط به کلاس Poor Dilation با F1-score برابر با 0.8649 بود، در حالی که ضعیف‌ترین عملکرد در کلاس Moderate مشاهده شد که تنها به F1-score معادل 0.3590 رسید.

به طور کلی، نتایج نشان می‌دهد که اگرچه EfficientNet-B1 نسبت به B0 توانست بهبود اندکی در برخی کلاس‌ها ایجاد کند، اما همچنان در تفکیک دقیق‌تر برخی از کلاس‌های میانی مانند Moderate و High ضعف نشان داد. این موضوع اهمیت افزایش حجم داده‌ها و طراحی استراتژی‌های بهبود آموزش را در مدل‌های بزرگ‌تر خانواده EfficientNet برجسته می‌سازد.

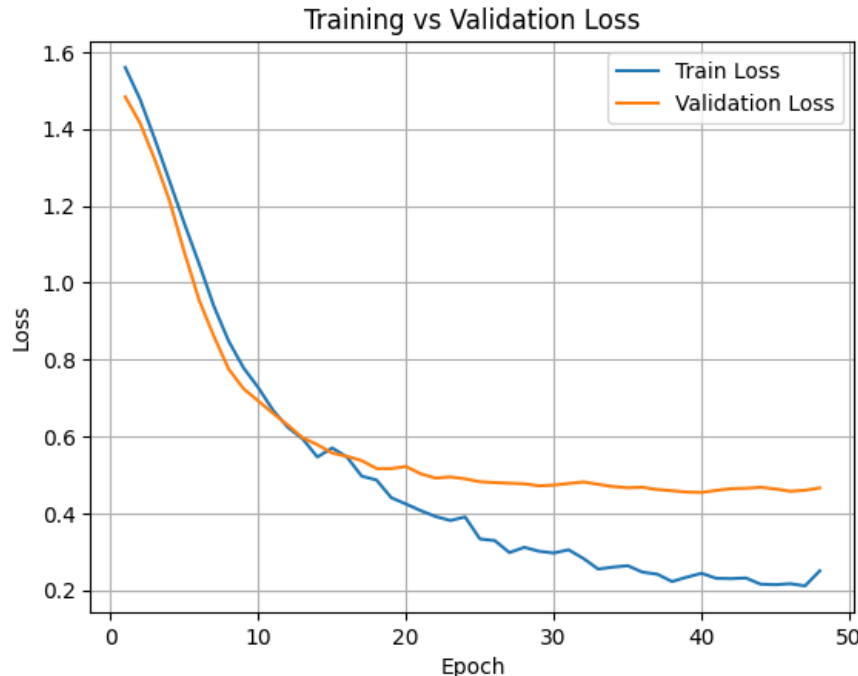
EfficientNet-B2 2.4.5

EfficientNet-B2 نسخه‌ی پیشرفته‌تر خانواده EfficientNet است که نسبت به B0 و B1 از ظرفیت بیشتری برای یادگیری ویژگی‌های پیچیده برخوردار است. این مدل با افزایش همزمان عمق، عرض شبکه و رزولوشن ورودی، قادر است جزئیات ریزتری از تصاویر را استخراج کرده و در دسته‌بندی‌های پیچیده‌تر عملکرد بهتری نشان دهد. به همین دلیل، EfficientNet-B2 در بسیاری از کاربردهای پردازش تصویر پزشکی به‌عنوان مدلی متوازن میان دقت و حجم محاسبات شناخته می‌شود.

در این پروژه، مدل B2 با وزن‌های از پیش‌آموزش‌دیده بر روی ImageNet بارگذاری و لایه‌ی خروجی آن برای پنج کلاس اصلی جایگزین شد. همانند مراحل قبلی، فرآیند آموزش شامل استفاده از تکنیک‌های کاهش نرخ یادگیری (ReduceLROnPlateau)، توقف زودهنگام (Early Stopping) و افزایش داده (Augmentation) بود. هدف از به‌کارگیری B2، بررسی توانایی آن در مقایسه با نسخه‌های کوچک‌تر و ارزیابی پتانسیل آن برای دستیابی به نتایج قابل اتکا در زمینه تشخیص شدت کاتاراکت بود.

نمودار آموزش و اعتبارسنجی نشان می‌دهد که هر دو منحنی روند نزولی پایداری داشتند و فاصله میان

Train Loss و Validation Loss در دوره‌های پایانی به حداقل رسید. این موضوع بیانگر آن است که مدل توانسته است به خوبی تعمیم پیدا کند و دچار بیش‌برازش قابل توجه نشده است.



شکل 18 EFFICIENTNET-B2 Loss per Epoch

نتایج نهایی این مدل بسیار امیدوارکننده بود. EfficientNet-B2 به دقت کلی (Accuracy) برابر با 83.07 درصد دست یافت. همچنین مقادیر Macro F1 و Weighted F1 هر دو برابر با حدود 0.8239 و 0.8302

گزارش شدند. این ارقام نشان می‌دهند که مدل در تمام کلاس‌ها عملکردی متوازن و نسبتاً قوی داشته است. بهترین عملکرد مربوط به کلاس Low nuclear density با F1-score معادل 0.8830 بود، در حالی که ضعیف‌ترین کلاس، Moderate با F1-score برابر با 0.7556 محسوب می‌شود که با وجود این، همچنان نسبت به مدل‌های پیشین بهبود چشمگیری نشان داده است.

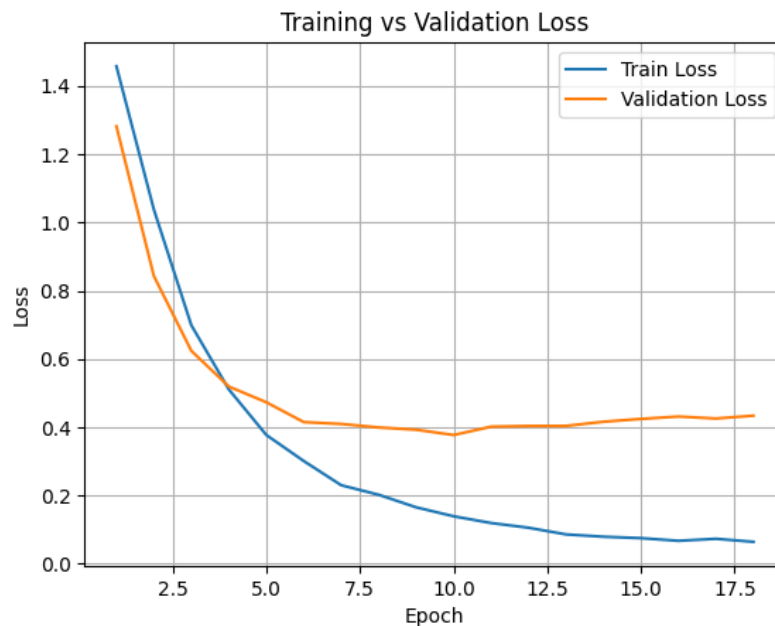
به طور کلی، EfficientNet-B2 در این پروژه بهترین نتایج را در مقایسه با مدل‌های قبلی خانواده EfficientNet و همچنین برخی معماری‌های ResNet نشان داد. این امر بیانگر تناسب بالای ظرفیت مدل با حجم داده و میزان پیچیدگی مسئله است و انتخاب آن به‌عنوان یکی از مدل‌های اصلی برای ادامه کار منطقی و قابل دفاع به نظر می‌رسد.

EfficientNet-B3 2.4.6

EfficientNet-B3 نسخه‌ی پیشرفته‌تری از این خانواده است که نسبت به B2 عمق و ظرفیت بالاتری دارد. این مدل با افزایش تعداد لایه‌ها و پارامترها، توانایی بیشتری در یادگیری ویژگی‌های ظریف و پیچیده تصاویر نشان می‌دهد. در حوزه‌هایی مانند پردازش تصویر پزشکی، که نیازمند شناسایی تفاوت‌های بسیار جزئی میان کلاس‌ها است، معماری‌هایی با ظرفیت بیشتر می‌توانند مزایای قابل توجهی ارائه کنند.

در این پروژه، EfficientNet-B3 نیز با وزن‌های از پیش آموزش‌دیده بارگذاری شد و لایه‌ی خروجی آن برای پنج کلاس اصلی تنظیم گردید. فرآیند آموزش مشابه مدل‌های قبلی شامل استفاده از تکنیک‌هایی همچون کاهش نرخ یادگیری پویا، توقف زودهنگام، و افزایش داده‌های متنوع بود. هدف اصلی از آزمایش B3، بررسی میزان بهبود عملکرد نسبت به B2 و ارزیابی کارایی آن در تفکیک دقیق‌تر مراحل مختلف کاتاراکت بود.

نتایج حاصل از ارزیابی نشان داد که EfficientNet-B3 بالاترین سطح عملکرد را در مقایسه با تمامی مدل‌های قبلی ارائه داده است. این مدل به دقت کلی (Accuracy) برابر با 92.03 درصد دست یافت. همچنین مقادیر Macro F1 برابر با 0.9116 و Weighted F1 برابر با 0.9191 گزارش شدند که نشان‌دهنده‌ی تعادل بسیار مناسب میان تمامی کلاس‌ها است. عملکرد کلاس‌ها نیز بسیار قابل توجه بود: کلاس Low nuclear density با F1-score برابر با 0.9583 بهترین نتیجه را داشت، و حتی کلاس‌های پیچیده‌تر مانند Moderate و High نیز توانستند به ترتیب به F1-score برابر با 0.8438 و 0.8750 برسند. نمودار Loss بر حسب Epoch نیز به صورت زیر می‌باشد:



شکل 19 EFFICIENTNET-B3 Loss per Epoch

این نتایج بیانگر آن است که EfficientNet-B3 نه تنها توانایی بالایی در استخراج ویژگی‌های دقیق از تصاویر جراحی کاتاراکت دارد، بلکه در تمامی کلاس‌ها تعادل و ثبات عملکرد را حفظ کرده است. به همین دلیل، می‌توان B3 را به عنوان قدرتمندترین مدل مورد آزمایش در این پروژه معرفی کرد.

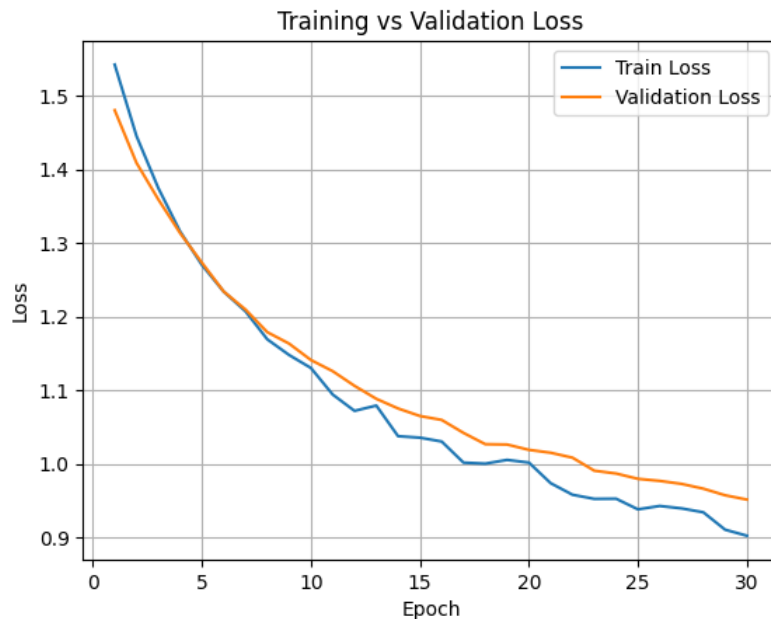
EfficientNet-B4 2.4.7

EfficientNet-B4 یکی از نسخه‌های بزرگ‌تر و قدرتمندتر خانواده EfficientNet است که نسبت به B0 تا B3 دارای عمق و رزولوشن بالاتری است. این افزایش ظرفیت به مدل اجازه می‌دهد تا ویژگی‌های بسیار پیچیده‌تری را از تصاویر استخراج کند، اما در عین حال نیازمند داده‌های بیشتر و سخت‌گیرانه‌تر برای آموزش مناسب است. به همین دلیل، اگر مجموعه داده از نظر حجم یا تنوع محدود باشد، مدل‌های بزرگ‌تر این خانواده ممکن است دچار بیش‌برازش شوند یا کارایی آن‌ها کاهش یابد.

در این پروژه، EfficientNet-B4 نیز با وزن‌های از پیش آموزش دیده روی ImageNet بارگذاری شد و خروجی آن برای پنج کلاس مورد نظر بازطراحی گردید. آموزش این مدل همانند سایر نسخه‌ها با استفاده از تکنیک‌هایی نظیر کاهش پویا نرخ یادگیری، توقف زودهنگام و اعمال افزایش داده (Augmentation) انجام شد. هدف از استفاده از این مدل بررسی اثر افزایش مقیاس شبکه بر نتایج بود.

نمودار آموزش و اعتبارسنجی نشان داد که هرچند منحنی‌های Train Loss و Validation Loss روند نزولی داشتند و به ثبات نسبی رسیدند، اما فاصله میان آن‌ها در طول آموزش نسبتاً ثابت باقی ماند. این موضوع نشان

می‌دهد که مدل B4 به‌خوبی قادر به تعمیم‌دهی کامل بر روی داده‌ها نبوده و بخشی از ظرفیت بالای آن بلااستفاده مانده است.



شکل 20 EFFICIENTNET-B4 Loss per Epoch

از نظر نتایج عددی، مدل EfficientNet-B4 به دقت کلی (Accuracy) برابر با 64.81 درصد دست یافت. همچنین Macro F1 برابر با 0.6399 و Weighted F1 برابر با 0.6383 گزارش شد. بهترین عملکرد مربوط به کلاس Poor Dilation با F1-score برابر با 0.8052 بود، در حالی که ضعیف‌ترین عملکرد در کلاس High مشاهده شد که تنها به F1-score برابر با 0.4000 رسید.

این نتایج نشان می‌دهد که علی‌رغم ظرفیت بالای معماری B4، در این پروژه مدل نتوانست عملکردی هم‌تراز با نسخه‌های سبک‌تر (مانند B2 و B3) ارائه دهد. علت اصلی این مسئله احتمالاً محدودیت حجم داده‌ها و نیاز این معماری به تعداد نمونه‌های بسیار بیشتر برای یادگیری مؤثر است. بنابراین، اگرچه EfficientNet-B4 از نظر تئوری قدرتمندتر است، اما در عمل برای چنین مجموعه داده‌ای بهینه‌ترین انتخاب محسوب نمی‌شود.

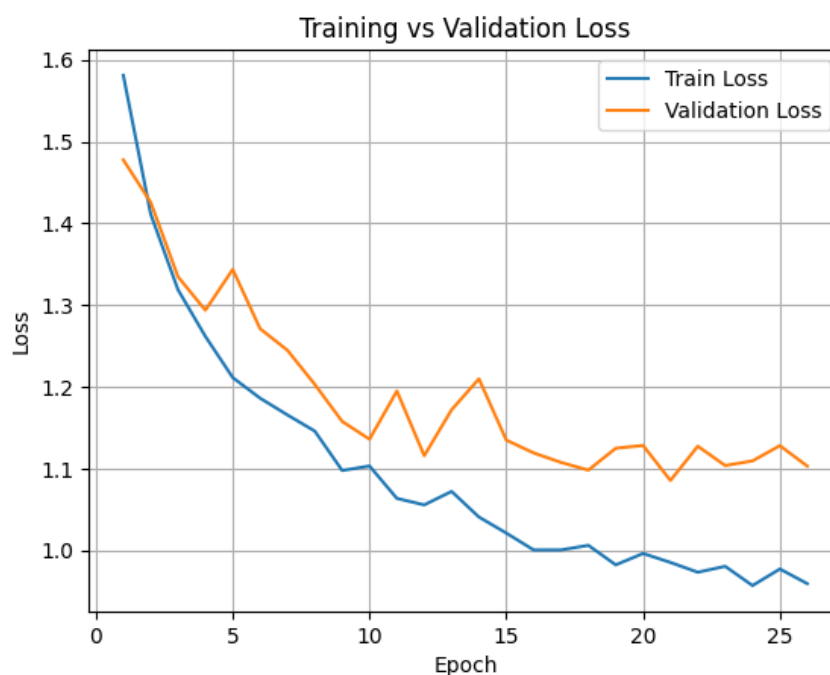
2.4.8 DenseNet121

DenseNet121 یکی از معماری‌های پیشرفته‌ی شبکه‌های عصبی کانولوشنی است که با معرفی اتصالات متراکم (Dense Connections) تحول مهمی در حوزه یادگیری عمیق ایجاد کرد. در این معماری، هر لایه نه تنها به خروجی لایه قبل، بلکه به تمامی لایه‌های پیشین متصل است. این ساختار باعث می‌شود ویژگی‌های استخراج‌شده در سطوح مختلف شبکه به‌طور کامل به اشتراک گذاشته شوند و یادگیری بهینه‌تر و کارآمدتری صورت گیرد. از مهم‌ترین مزایای DenseNet می‌توان به کاهش تعداد پارامترها، بهبود جریان گرادینانها و جلوگیری از مشکل

ناپدید شدن گرادیان اشاره کرد.

در این پروژه، DenseNet121 با وزن‌های از پیش‌آموزش‌دیده بر روی دیتاست ImageNet بارگذاری و لایه‌ی خروجی آن برای پنج کلاس اصلی جایگزین شد. هدف از به‌کارگیری این مدل، بررسی عملکرد معماری DenseNet در مقایسه با مدل‌های خانواده ResNet و EfficientNet بود. فرآیند آموزش شامل استفاده از تکنیک‌های استاندارد مانند کاهش نرخ یادگیری پویا، توقف زودهنگام و افزایش داده‌ها بود.

نمودار آموزش و اعتبارسنجی نشان داد که هرچند خطای آموزش کاهش یافت و به ثبات نسبی رسید، اما فاصله‌ی میان Train Loss و Validation Loss به طور محسوسی باقی ماند. این موضوع می‌تواند نشانه‌ی بیش‌برازش یا نیاز مدل به داده‌های متنوع‌تر باشد.



شکل 21 DENSENET121 Loss per Epoch

نتایج نهایی نشان داد که DenseNet121 به دقت کلی (Accuracy) برابر با 55.03 درصد دست یافت. همچنین مقادیر Macro F1 برابر با 0.5270 و Weighted F1 برابر با 0.5252 گزارش شدند. بهترین عملکرد مدل مربوط به کلاس Poor Dilation با F1-score معادل 0.7059 بود، در حالی که ضعیف‌ترین عملکرد در کلاس High با F1-score تنها برابر با 0.3462 مشاهده شد.

این نتایج بیانگر آن است که اگرچه DenseNet121 از نظر معماری بسیار قدرتمند و کارآمد است، اما در این پروژه و با توجه به حجم و تنوع محدود داده‌ها نتوانست عملکردی هم‌سطح با EfficientNet-B2 و B3 ارائه دهد. به نظر می‌رسد که برای بهره‌گیری کامل از ظرفیت DenseNet، نیاز به مجموعه‌داده‌ی بزرگ‌تر و متنوع‌تر وجود دارد.

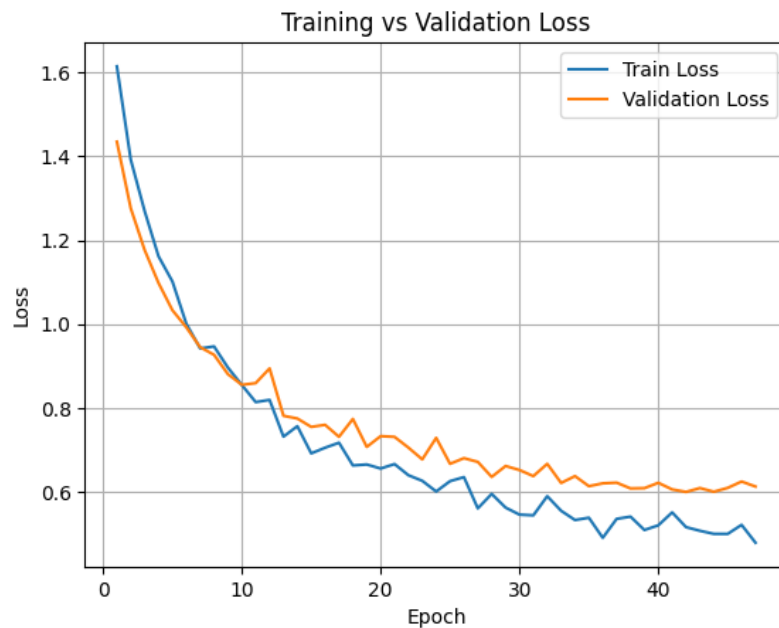
DenseNet169 2.4.9

DenseNet169 نسخه‌ی بزرگ‌تر معماری DenseNet است که با افزایش تعداد لایه‌ها نسبت به DenseNet121، ظرفیت بالاتری برای استخراج ویژگی‌های پیچیده دارد. همانند سایر نسخه‌های این خانواده، ویژگی اصلی آن وجود اتصالات متراکم (Dense Connections) است که باعث می‌شود تمامی لایه‌ها به خروجی لایه‌های قبلی دسترسی داشته باشند. این موضوع موجب استفاده‌ی بهینه‌تر از ویژگی‌ها، کاهش نیاز به پارامترهای اضافی، و بهبود جریان گرادیان‌ها در طول آموزش می‌شود.

در این پروژه، DenseNet169 با وزن‌های از پیش‌آموزش‌دیده بارگذاری شد و لایه‌ی نهایی آن برای پنج کلاس اصلی بازطراحی گردید. فرآیند آموزش مشابه سایر مدل‌ها شامل استفاده از افزایش داده‌ها، کاهش نرخ یادگیری پویا، و توقف زودهنگام بود. هدف از به‌کارگیری DenseNet169، بررسی توانایی نسخه‌ی عمیق‌تر این خانواده در بهبود عملکرد نسبت به DenseNet121 بود.

نتایج به‌دست‌آمده نشان داد که DenseNet169 به‌طور چشمگیری بهتر از نسخه‌ی 121 عمل کرد. این مدل توانست به دقت کلی (Accuracy) برابر با 83.13 درصد دست یابد. همچنین مقادیر Macro F1 برابر با 0.8212 و Weighted F1 برابر با 0.8294 گزارش شدند که نشان‌دهنده‌ی عملکردی متوازن و پایدار میان کلاس‌های مختلف است. بهترین عملکرد مربوط به کلاس Poor Dilation با F1-score برابر با 0.8980 بود، و کلاس‌های Low و Mature نیز به ترتیب با F1-score معادل 0.8817 و 0.8732 نتایج بسیار خوبی داشتند. در مقابل، ضعیف‌ترین عملکرد در کلاس Moderate مشاهده شد که به F1-score برابر با 0.7111 رسید.

نمودار آموزش و اعتبارسنجی نشان داد که هر دو منحنی Train Loss و Validation Loss روند نزولی پایداری داشتند و به همگرایی خوبی رسیدند. این مسئله نشان می‌دهد که مدل DenseNet169 توانسته است تعادل مناسبی میان آموزش و تعمیم‌پذیری حفظ کند و نسبت به DenseNet121 پیشرفت قابل توجهی داشته باشد.



شکل 22 DENSENET169 Loss per Epoch

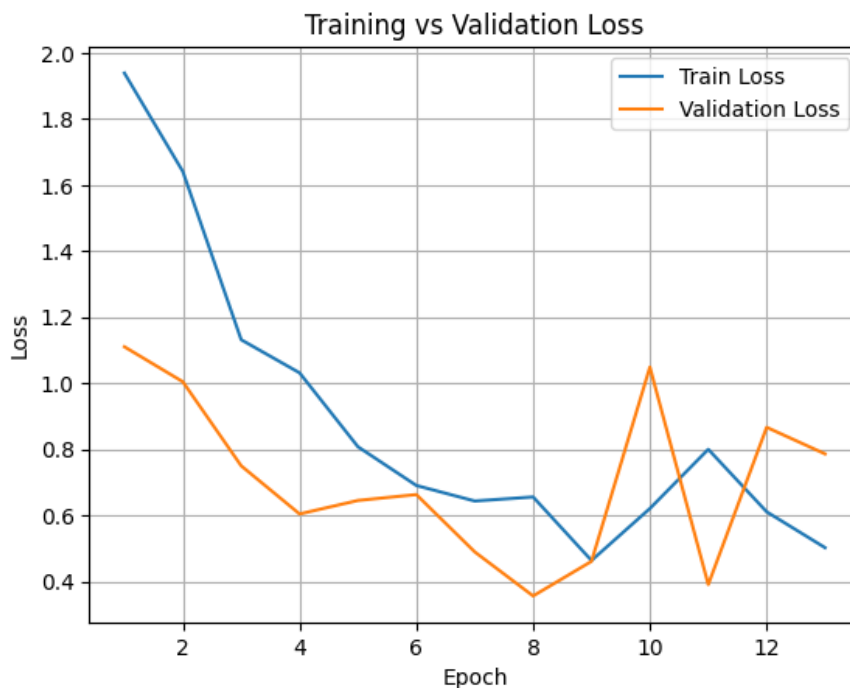
به طور کلی، DenseNet169 در این پروژه به یکی از موفق‌ترین مدل‌ها تبدیل شد و از نظر دقت و تعادل میان کلاس‌ها توانست عملکردی هم‌سطح EfficientNet-B2 و B3 ارائه دهد.

VGG16 2.4.10

VGG16 یکی از معماری‌های کلاسیک و بسیار تأثیرگذار در حوزه شبکه‌های عصبی کانولوشنی است که توسط گروه Visual Geometry Group دانشگاه آکسفورد معرفی شد. این مدل با ۱۶ لایه شامل لایه‌های کانولوشنی و Fully Connected طراحی شده و به دلیل ساختار ساده اما عمیق، در سال‌های ابتدایی توسعه شبکه‌های عمیق به عنوان یکی از معماری‌های استاندارد شناخته می‌شد. ایده اصلی در VGG16 استفاده از فیلترهای کوچک (۳×۳) در چندین لایه متوالی است که امکان استخراج تدریجی ویژگی‌ها را فراهم می‌کند.

در این پروژه، VGG16 با وزن‌های از پیش آموزش‌دیده بر روی دیتاست ImageNet بارگذاری شد و لایه خروجی آن برای پنج کلاس اصلی بازطراحی گردید. هدف از استفاده از این معماری، ارزیابی عملکرد یکی از مدل‌های کلاسیک در مقایسه با معماری‌های جدیدتر مانند ResNet، DenseNet و EfficientNet بود. فرآیند آموزش شامل افزایش داده‌ها، تنظیم پویا نرخ یادگیری و توقف زودهنگام بود.

نمودار آموزش و اعتبارسنجی نشان می‌دهد که مدل VGG16 در طول دوره‌ها نوسانات بیشتری نسبت به سایر معماری‌ها داشته است. منحنی Validation Loss افت و خیزهای متعددی داشت که می‌تواند نشان‌دهنده حساسیت بالای این مدل به پارامترهای آموزشی و کمبود داده‌های متنوع باشد. با این وجود، مدل توانست به سطح قابل قبولی از همگرایی برسد.



شکل 23 VGG16 Loss per Epoch

از نظر عملکرد عددی، VGG16 به دقت کلی (Accuracy) برابر با 78.75 درصد دست یافت. همچنین مقادیر Macro F1 برابر با 0.7848 و Weighted F1 برابر با 0.7902 گزارش شد. بهترین عملکرد مربوط به کلاس Mature با F1-score معادل 0.9615 بود، در حالی که ضعیف‌ترین عملکرد در کلاس Poor Dilation با F1-score برابر با 0.6923 مشاهده شد. این نتایج نشان می‌دهند که مدل در تفکیک کلاس‌هایی با ویژگی‌های واضح‌تر (مانند Poor Dilation) موفق‌تر عمل کرده است.

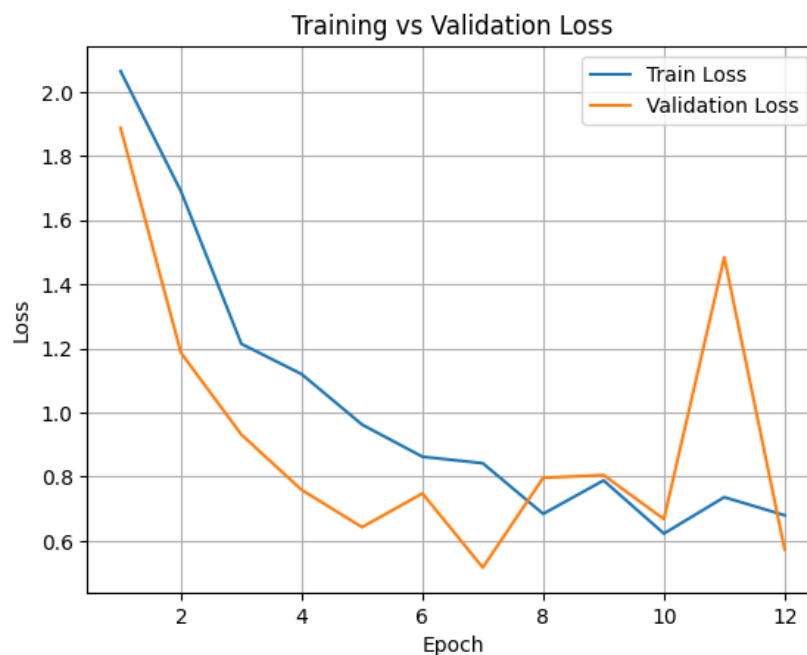
به طور کلی، هرچند VGG16 یکی از معماری‌های قدیمی‌تر است و از نظر کارایی محاسباتی نسبت به مدل‌های جدیدتر محدودیت‌هایی دارد، اما در این پروژه عملکرد نسبتاً مطلوبی ارائه داد و نشان داد که همچنان می‌تواند در مسائل خاص با داده‌های مناسب نتایج خوبی به همراه داشته باشد.

VGG19 2.4.11

VGG19 یکی از نسخه‌های عمیق‌تر معماری VGG است که با ۱۹ لایه طراحی شده و نسبت به VGG16 ظرفیت بیشتری برای استخراج ویژگی‌های پیچیده دارد. ایده اصلی در این معماری نیز همانند VGG16 بر پایه استفاده متوالی از فیلترهای کوچک 3×3 و لایه‌های کانولوشنی پشت سر هم بنا شده است. عمق بیشتر در VGG19 امکان استخراج الگوهای دقیق‌تر را فراهم می‌کند، اما این افزایش عمق همچنین به هزینه محاسباتی بالاتر و احتمال بیش‌برازش منجر می‌شود، به‌ویژه زمانی که حجم داده محدود باشد.

در این پروژه، مدل VGG19 با وزن‌های از پیش‌آموزش‌دیده بر روی ImageNet بارگذاری شد و لایه خروجی آن برای پنج کلاس اصلی جایگزین گردید. فرآیند آموزش مشابه سایر مدل‌ها شامل استفاده از افزایش داده‌ها، کاهش نرخ یادگیری پویا، و توقف زودهنگام بود. هدف از به‌کارگیری VGG19 بررسی تأثیر افزایش عمق شبکه نسبت به VGG16 بر عملکرد نهایی بود.

نتایج آموزش نشان داد که مدل روند همگرایی مناسبی داشت، هرچند که منحنی Validation Loss در برخی نقاط دچار نوسان‌های شدیدی شد. این ناپایداری می‌تواند ناشی از حساسیت بالای VGG19 به داده‌های محدود و نیاز آن به تنظیمات دقیق‌تر ابرپارامترها باشد.



شکل 24 VGG19 Loss per Epoch

از نظر عملکرد عددی، VGG19 به دقت کلی (Accuracy) برابر با 81.87 درصد دست یافت. همچنین مقادیر Macro F1 برابر با 0.8085 و Weighted F1 برابر با 0.8171 گزارش شد. بهترین عملکرد مربوط به کلاس Low nuclear density با F1-score معادل 0.9157 بود، در حالی که ضعیف‌ترین عملکرد در کلاس High مشاهده شد که تنها به F1-score برابر با 0.6667 رسید. در کلاس Poor Dilation نیز عملکرد مدل

بسیار خوب بود و توانست به F1-score معادل 0.8438 برسد.

به طور کلی، نتایج نشان داد که VGG19 در مقایسه با VGG16 اندکی بهبود در برخی کلاس‌ها ایجاد کرده است، اما هزینه محاسباتی بیشتر و ناپایداری در منحنی‌های اعتبارسنجی از محدودیت‌های آن محسوب می‌شود. این مسئله نشان می‌دهد که اگرچه مدل‌های VGG همچنان قادر به ارائه نتایج قابل قبول هستند، اما معماری‌های جدیدتر مانند EfficientNet و DenseNet کارایی بهتری در داده‌های پزشکی پیچیده ارائه می‌دهند.

2.4.12 جمع‌بندی

در این بخش، مجموعه‌ای از معماری‌های مختلف شبکه‌های عصبی کانولوشنی شامل خانواده‌های ResNet، EfficientNet، DenseNet و VGG بر روی داده‌های استخراج‌شده از مراحل جراحی کاتاراکت مورد ارزیابی قرار گرفتند. هدف اصلی، شناسایی مدلی بود که علاوه بر دقت بالا، تعادل مناسبی میان کلاس‌ها برقرار کرده و توانایی تعمیم‌پذیری قابل قبولی داشته باشد.

مدل‌های ResNet18 و ResNet50 نتایج متوسطی ارائه دادند؛ ResNet18 با دقت حدود 75٪ توانست الگوهای اولیه را به خوبی بیاموزد، در حالی که ResNet50 با وجود عمق بیشتر، تنها اندکی بهبود یافت. این نتایج نشان داد که افزایش عمق در غیاب داده‌های بسیار بزرگ، لزوماً به بهبود قابل توجه منجر نمی‌شود.

در خانواده EfficientNet، نسخه‌های پایه‌تر B0 و B1 عملکرد نسبتاً ضعیفی داشتند و به دلیل محدودیت ظرفیت، نتوانستند تفاوت‌های پیچیده میان کلاس‌ها را به خوبی تشخیص دهند. در مقابل، EfficientNet-B2 و به‌ویژه EfficientNet-B3 بهترین عملکرد را در کل آزمایش‌ها به نمایش گذاشتند؛ B2 با دقت حدود 83٪ و B3 با دقت بیش از 92٪ توانستند تمامی کلاس‌ها را با ثبات و تعادل بالا دسته‌بندی کنند. نسخه‌های بزرگ‌تر مانند B4، به دلیل نیاز به داده‌های حجیم‌تر، نتایج ضعیف‌تری داشتند و نشان دادند که مقیاس بیش از حد در این پروژه سودمند نبوده است.

مدل‌های DenseNet نیز کارایی جالبی نشان دادند. DenseNet121 با دقت حدود 55٪ نتایج نسبتاً ضعیفی داشت، اما DenseNet169 توانست عملکردی مشابه EfficientNet-B2 ارائه دهد و با دقت بیش از 83٪ و F1 متوازن، خود را به عنوان یکی از مدل‌های موفق این پروژه مطرح کند. این امر نشان‌دهنده اهمیت افزایش عمق در DenseNet و استفاده مؤثر از اتصالات متراکم برای استخراج بهتر ویژگی‌ها بود.

در نهایت، مدل‌های کلاسیک VGG16 و VGG19 نیز مورد بررسی قرار گرفتند. اگرچه این مدل‌ها از نظر معماری قدیمی‌تر هستند و هزینه محاسباتی بالاتری دارند، اما هر دو توانستند نتایج قابل قبولی ارائه دهند؛ VGG16 با دقت حدود 79٪ و VGG19 با دقت بیش از 81٪ نشان دادند که همچنان در مسائل پزشکی می‌توانند رقابتی عمل کنند، هرچند که نسبت به معماری‌های مدرن‌تر، ضعف بیشتری در تعمیم‌پذیری و پایداری دارند.

به طور کلی، نتایج این بخش نشان داد که بهترین مدل‌های پروژه EfficientNet-B3 و DenseNet169 بودند که هر دو عملکردی قوی، متوازن و قابل اعتماد ارائه دادند. این نتایج اهمیت انتخاب معماری مناسب متناسب با اندازه و کیفیت داده را نشان می‌دهد. در حالی که مدل‌های ساده‌تر توانستند شروع خوبی فراهم کنند، تنها مدل‌های با ظرفیت بهینه (نه خیلی کوچک و نه بیش از حد بزرگ) توانستند به نتایج برجسته برسند.

- Ambreen Gul, Sairam Ahmed, Samana Ali, and Ali Raza, "Phacoemulsification in senile white mature cataracts," *Pak. J. Ophthalmol.*, vol. 35, no. 4, pp. 265–270, 2019.
- J. Crispim, L. S. Jung, L. Paz, N. Allemann, and P. Schor, "The surgical challenges dense brunescant cataracts present," *Expert Rev. Ophthalmol.*, vol. 10, no. 1, pp. 13–22, 2014.
- S. S. Ermis, F. Öztürk, and Ü. Ü. Inan, "Comparing the efficacy and safety of phacoemulsification in white mature and other types of senile cataracts," *Br. J. Ophthalmol.*, vol. 87, no. 11, pp. 1356–1359, 2003.
- M. Gupta, J. Ram, A. Jain, J. Sukhija, and M. Chaudhary, "Correlation of nuclear density using the Lens Opacity Classification System III versus Scheimpflug imaging with phacoemulsification parameters," *J. Cataract Refract. Surg.*, vol. 39, no. 12, pp. 1818–1823, 2013.
- Z. Bencic, V. Zoric-Geber, Z. Vatauvuk, M. Ivekovic, and V. Mandic, "Comparison of ultrasound energy and corneal endothelial cell loss during phacoemulsification of mature and immature cataracts," *Coll. Antropol.*, vol. 31, no. 1, pp. 103–107, 2007.
- S. A. Lim, J. Hwang, K.-Y. Hwang, and S.-H. Chung, "Objective assessment of nuclear cataract: Comparison of double-pass and Scheimpflug systems," *J. Cataract Refract. Surg.*, vol. 40, no. 5, pp. 716–721, May 2014, doi: 10.1016/j.jcrs.2013.10.032.
- Z. Qiao, Q. Zhang, Y. Dong, and J.-J. Yang, "Application of SVM based on genetic algorithm in classification of cataract fundus images," in *Proc. IEEE Int. Conf. Imaging Systems and Techniques (IST)*, 2017, pp. 1–5 .
- V.-V. Nguyen and C.-L. Lin, "Enhancing cataract detection through hybrid CNN approach and image quadrature: A solution for precise diagnosis and improved patient care," *Electronics*, vol. 13, no. 2344, pp. 1–28, 2024, doi: 10.3390/electronics13122344 [16:2.pdf] .
- M. Yang, J.-J. Yang, Q. Zhang, Y. Niu, and J. Li, "Classification of retinal image for automatic cataract detection," in *Proc. IEEE 15th Int. Conf. e-Health Networking, Applications and Services (Healthcom)*, 2013, pp. 117–121 .
- **L. Woodard**, "The stability of the red reflex produced by different surgical ophthalmic microscopes," *Ophthalmology and Therapy*, vol. 10, no. 2, pp. 389–391, Jul. 2021, doi: 10.1007/s40123-021-00367-w.