



دانشگاه بوعلی سینا

دانشکده مهندسی

گروه مهندسی کامپیوتر

پایان نامه برای دریافت درجه کارشناسی در رشته مهندسی کامپیوتر

عنوان:

تحلیل احساس چهره با استفاده از شبکه عصبی عمیق

استاد راهنما:

دکتر حسن ختن‌لو

نگارش:

علیرضا دلاوری

اسفند ۱۴۰۱



دانشگاه بوعلی سینا

مشخصات پایان نامه تحصیلی

عنوان:

تحلیل احساس چهره با استفاده از شبکه عصبی عمیق

نام نویسنده: علیرضا دلاوری

نام استاد راهنما: حسن ختن لو

دانشکده: مهندسی

گروه آموزشی: کامپیوتر

رشته تحصیلی: مهندسی کامپیوتر

تاریخ دفاع: اسفند ۱۴۰۱

مقطع تحصیلی: کارشناسی

چکیده:

توانایی تعامل عاطفی یکی از اصلی ترین مسائل در تعیین نوع و اندازه تعاملات بین ماشین و انسان است و شناخت حالت چهره یک مهارت کلیدی در پیشرفت این تعاملات محسوب می شود. تشخیص حالت چهره با دقت بالا، به دلیل وجود پیچیدگی و تغییرپذیری حالت صورت و عوامل گوناگونی که بر آن تاثیر می گذارند، به سختی به دست می آید.

در این پژوهش، مطلوب تشخیص هفت احساس پایه (عصبانیت، انزجار، ترس، ناراحتی، تعجب، خنثی، شادی) با توجه به چهره افراد با استفاده از شبکه های عصبی عمیق می باشد. برای این منظور از مجموعه داده FER-2013 که از تصاویر طبیعی جمع آوری شده و عواملی همچون جنسیت، سن، قومیت و ... در آن در نظر گرفته شده است استفاده کرده ایم.

سه نوع مدل مبتنی بر شبکه عصبی ResNet34 و دو نوع از انواع EfficientNet که از معتبر ترین شبکه های عصبی می باشند ساخته و آن ها را بر روی مجموعه داده مذکور آزمایش کرده ایم. در نهایت دقت های ۶۳٫۹۷، ۶۵٫۷۱، ۶۴٫۷۳، ۶۴٫۳۲ و ۶۴٫۱۸ که نسبت به پیچیدگی مجموعه داده دقت بالایی محسوب می شوند را حاصل داده اند.

کلمات کلیدی: هوش مصنوعی، شبکه های عصبی عمیق، پردازش تصویر، CNN

کلیه‌ی امتیازهای این پایان‌نامه به دانشگاه بوعلی‌سینا تعلق دارد. در صورت استفاده از تمام یا بخشی از مطالب این پایان‌نامه در مجلات، کنفرانس‌ها و یا سخنرانی‌ها باید نام دانشگاه بوعلی‌سینا یا استاد راهنمای پایان‌نامه و نام دانشجو با ذکر مأخذ و ضمن کسب مجوز کتبی از دفتر دانشگاه ثبت شود. در غیر این صورت مورد پیگرد قانونی قرار خواهد گرفت.

از استاد بزرگوarm دکتر ختن‌لو برای تمام زحمات بی دریغ‌شان سپاسگزاری می‌کنم که بدون حمایت‌های ایشان اتمام این پژوهش ممکن نبود.

همچنین از استادان گرانمایه، جناب آقای دکتر منصوری‌زاده و خانم مهندس بطحائیان که زحمت داوری این پایان نامه را بر عهده داشتند نهایت تشکر را دارم.

## فهرست مطالب

۶	فهرست اشکال
۷	فهرست جداول
۸	۱- مقدمه
۸	۱-۱ کلیات
۸	۲-۱ یادگیری عمیق
۹	۳-۱ شبکه های عصبی کانولوشنی
۹	۴-۱ دیتاست
۱۱	۵-۱ معماری ResNet34
۱۱	۲- پیشینه پژوهش
۱۳	۳- روش پیشنهادی
۱۳	۳-۱ پیش پردازش
۱۳	۳-۲ مدل ابتدایی ResNet34
۱۵	۳-۳ مدل پیشنهادی اول
۱۶	۳-۴ مدل پیشنهادی دوم
۱۶	۳-۵ EfficientNet
۱۷	۴- پیاده سازی مدل ها و تحلیل و ارزیابی نتایج
۱۷	۴-۱ مقدمه
۱۷	۴-۲ مدل ابتدایی ResNet34
۱۹	۴-۳ مدل پیشنهادی اول
۲۲	۴-۴ مدل پیشنهادی دوم
۲۴	۴-۵ EfficientNet-B3
۲۷	۴-۶ EfficientNet-B5
۲۹	۴-۷ ارزیابی مدل ها
۳۱	۵- نتیجه گیری و پیشنهادها
۳۱	۵-۱ مقدمه
۳۱	۵-۲ بررسی نتایج پژوهش
۳۱	۵-۳ پیشنهادها جهت مطالعات آتی
۳۲	منابع

## فهرست اشکال

شکل ۱- تصاویر نمونه از دیتاست FER-۲۰۱۳ به همراه برچسب متناظر.....	۱۱
شکل ۲- اتصال میانبر در معماری RESNET۳۴.....	۱۴
شکل ۳- معماری شبکه عصبی RESNET۳۴.....	۱۴
شکل ۴- ماتریس درهم‌ریختگی RESNET۳۴ بر روی داده‌های آموزشی (TRAIN).....	۱۸
شکل ۵- ماتریس درهم‌ریختگی RESNET۳۴ بر روی داده‌های آزمایشی (VALIDATION).....	۱۸
شکل ۶- نمودار ROC مدل RESNET۳۴ بر روی داده‌های آموزشی.....	۱۸
شکل ۷- نمودار ROC مدل RESNET۳۴ بر روی داده‌های آزمایشی.....	۱۹
شکل ۸- ماتریس درهم‌ریختگی مدل پیشنهادی اول بر روی داده‌های آزمایشی (VALIDATION).....	۲۰
شکل ۹- ماتریس درهم‌ریختگی مدل پیشنهادی اول بر روی داده‌های آموزشی (TRAIN).....	۲۰
شکل ۱۰- نمودار ROC مدل پیشنهادی اول بر روی داده‌های آموزشی (TRAIN).....	۲۱
شکل ۱۱- نمودار ROC مدل پیشنهادی اول بر روی داده‌های آزمایشی (VALIDATION).....	۲۱
شکل ۱۲- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آزمایشی (VALIDATION).....	۲۳
شکل ۱۳- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آموزشی (TRAIN).....	۲۳
شکل ۱۴- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آموزشی (TRAIN).....	۲۳
شکل ۱۵- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آزمایشی (VALIDATION).....	۲۴
شکل ۱۶- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آزمایشی (VALIDATION).....	۲۵
شکل ۱۷- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آموزشی (TRAIN).....	۲۵
شکل ۱۸- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آموزشی (TRAIN).....	۲۶
شکل ۱۹- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آزمایشی (VALIDATION).....	۲۶
شکل ۲۰- نمودار تغییرات خطا در دوره‌ها (سمت راست) و نمودار تغییرات دقت در دوره‌ها (سمت چپ).....	۲۷
شکل ۲۱- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آزمایشی (VALIDATION).....	۲۸
شکل ۲۲- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آزمایشی (VALIDATION).....	۲۸
شکل ۲۳- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آموزشی (TRAIN).....	۲۸
شکل ۲۴- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آموزشی (TRAIN).....	۲۹

## فهرست جداول

- جدول ۱- تعداد داده‌های هر کلاس در داخل دیتاست FER-۲۰۱۳ ..... ۱۰
- جدول ۲- داده افزایشی‌های انجام شده بر روی داده‌ها ..... ۱۳
- جدول ۳- ضریب‌های وزنی اعمال شده در هر یک از کلاس‌ها در دسته‌بندی پیشنهادی دوم ..... ۱۶
- جدول ۴- دقت مدل‌های مطرح شده بر روی داده‌های آموزشی و آزمایشی ..... ۲۹
- جدول ۵- AUC مدل‌های مطرح شده بر روی کلاس‌های دیتاست FER-۲۰۱۳ ..... ۳۰

# فصل اول

## ۱- مقدمه

### ۱-۱ کلیات

تجزیه و تحلیل بیان صورت یک حوزه تحقیقاتی چالش برانگیز در زمینه محاسبات عاطفی است. [1] محاسبه عاطفی سیستمی است که می‌تواند احساسات انسان را تشخیص دهد. احساسات انسانی در طول ارتباطات کلامی و غیر کلامی تا حد زیادی در چهره انسان منعکس می‌شود. به این ترتیب، به نظر می‌رسد که چهره انسان به احساسات بیشتر حساس است. برای معنادار کردن تعامل انسان و رایانه، ضروری است که رایانه‌ها احساسات انسان را به روشی که یک انسان تشخیص می‌دهد تشخیص دهند. [2] همچنین تشخیص احساسات انسانی از طریق هوش مصنوعی یکی از محبوب‌ترین زمینه‌های تحقیقاتی در بین محققان است و زمینه‌های تعامل رایانه انسانی (HCI) و محاسبه عاطفی به طور گسترده‌ای برای درک احساسات انسان استفاده می‌شود. انسان‌ها به طور کلی از وسایل غیر مستقیم و غیر کلامی زیادی برای انتقال احساسات خود استفاده می‌کنند. حالت چهره به جایگاه یا جابه‌جایی ماهیچه‌های صورت در زیر پوست مربوط است و می‌تواند انتقال دهنده، احساسات یک فرد برای بیننده باشد. حالت چهره خود نوعی ارتباط غیر کلامی است. حالت چهره از ابزارهای اولیه برقراری ارتباط در انسان‌ها است اما این پدیده در دیگر گونه‌ها هم دیده شده است.

### ۲-۱ یادگیری عمیق

یادگیری عمیق یک زیرشاخه از یادگیری ماشین می‌باشد که موفقیت‌های بسیاری در زمینه‌های مختلف کسب کرده است. با فرض این که داده ورودی یک تصویر باشد، روشهای یادگیری عمیق، پیکسل‌های خام تصویر را، از طریق یک مدل چندلایه و به صورت سلسله مراتبی، به مجموعه‌هایی از ویژگی‌ها تبدیل می‌کنند، که این ویژگی‌ها



در نهایت برای دسته‌بندی، به طبقه‌بند داده خواهد شد روش یادگیری عمیق در قالب معماری‌های مختلفی مانند شبکه‌های عصبی کانولوشنی (CNNs) پیاده‌سازی شده‌اند.

### ۳-۱ شبکه‌های عصبی کانولوشنی

شبکه‌های عصبی کانولوشنی، یکی از موفق‌ترین معماری‌ها در میان معماری‌های یادگیری عمیق هستند. شبکه‌های کانولوشنی عصبی، مدل‌های یادگیری عمیقی هستند که برداشتهای سطح میانی و سطح-بالا از داده خام بدست می‌آورند. یکی از مزیت‌های مهم شبکه‌های کانولوشنی این است که این دسته از شبکه‌ها، ویژگی‌ها را به طور مستقیم از داده ورودی استخراج می‌کنند و دیگر نیازی به طراحی یک استخراج کننده ویژگی مستقل در این شبکه‌ها وجود ندارد. دلیل نامگذاری این شبکه‌ها، استفاده از عملگر کانولوشن در ساختار آنها می‌باشد. در شبکه‌های عصبی کانولوشنی معمولاً چندین لایه کانولوشنی به کار گرفته می‌شود تا از این طریق یک برداشت سطح بالا از داده بدست بیاید. معمولاً بعد از هر لایه کانولوشنی یک لایه غیرخطی‌ساز و یک لایه ادغام قرار خواهد گرفت. لایه غیرخطی‌ساز کمک می‌کند تا برداشتهای سطح بالا بدست بیاید و لایه ادغام کمک می‌کند تا خروجی اصلاح شده و حساسیت نسبت به تغییرات نامطلوب کم شود. لایه ادغام، خلاصه ویژگی‌های آماری مقادیر خروجی همسایه‌های یک پیکسل ورودی را جایگزین خروجی آن پیکسل ورودی می‌کند. این مدل‌ها معمولاً داده‌های تصویری خام هستند (مثلاً پیکسل‌ها) و خروجی آنها می‌تواند بازنمایی سطح بالا یا خروجی‌های متناسب با اهداف طبقه‌بندی باشد. در مدل‌های کانولوشنی، بعد از چندین لایه کانولوشنی، معمولاً یک لایه تمام متصل برای اهداف طبقه‌بندی قرار داده می‌شود.

### ۴-۱ دیتاست

مجموعه داده‌ای مورد استفاده برای آموزش و ارزیابی مدل‌ها، مجموعه داده‌ی FER-2013 [3] می‌باشد. این مجموعه توسط پیر لوک کاریر و آرون کوریل ساخته شد. مجموعه داده با استفاده از Image search google جمع‌آوری شد و برای جمع‌آوری آنها ۱۸۹ کلمه‌ی مرتبط با احساسات، مانند ( blissful به معنی سرخوش) و

eurage به معنی خشمگین و ..... مورد جستجو قرار گرفت. این کلمات با کلماتی مرتبط با جنسیت، سن یا قومیت ترکیب شدند و در مجموعه ۶۰۰ عبارت برای جستجوی تصاویر صورت مورد استفاده قرار گرفت. ۱۰۰۰ عکس اول مرتبط با هر عبارت مورد جستجو، انتخاب شدند پردازش‌های بعدی روی آنها صورت گرفت. از سامانه‌ی تشخیص چهره‌ی کتابخانه‌ی opencv برای مشخص کردن چارچوب چهره استفاده شد و پس از برچسب‌گذاری‌های انسانی به اصلاح اندازه‌ی این چارچوب و حذف کردن عکس‌های تکراری پرداختند تصاویر تائید و اصلاح‌شده به اندازه‌ی ۴۸×۴۸ درآمدند و به عکس سیاه‌وسفید تبدیل شدند. مهدی میرزا و ایان گودفلو، زیرمجموعه‌ای از تصاویر را برای چالش موردنظر آماده کرده و آن در این ۷ حالت احساسی پایه طبقه‌بندی کردند [3]. مجموعه داده‌ی نهایی شامل ۳۵۸۸۷ تصویر است که تعداد آن در هر کلاس در جدول زیر آورده شده است.

جدول ۱- تعداد داده‌های هر کلاس در داخل دیتاست FER-2013

تعداد داده	کلاس احساسی
۴۹۵۳	عصبانی (Anger)
۵۴۷	انزجار (Disgust)
۵۱۲۱	ترس (Fear)
۸۹۸۹	خوشحالی (Happiness)
۶۰۷۷	ناراحتی (Sadness)
۴۰۰۲	تعجب (Surprise)
۶۱۹۸	خنثی (Neutral)

مجموعه داده‌ی FER-2013 به دلیل نحوه‌ی جمع‌آوری آن میتواند دارای خطای برچسب‌گذاری باشد. اما ایان گودفلو متوجه شد که دقت انسان روی این مجموعه داده در حدود ۵±۶۵٪ است [3].



شکل ۱- تصاویر نمونه از دیتاست FER-2013 به همراه برچسب متناظر

## ۵-۱ معماری ResNet34

عدد ۳۴ در resnet34 تعداد لایه‌های شبکه عصبی را نشان می‌دهد. Residual Unit block زیرساخت شبکه ResNet-34 است و کل شبکه عمدتاً از آن تشکیل شده است. حین یادگیری در هر Residual خطا کاهش پیدا میکند یا می‌توان گفت یادگیری بهبود مییابد. به همین دلایل به آن Residual Learning یا یادگیری افزایشی نیز گفته می‌شود. یک اتصال میانبر ( skip connection ) توسط هر بلوک برای رد شدن از لایه‌های کانولوشن استفاده شده است، که به طور موثر مشکل ناپدید شدن گرادیان ( gradient disappearance ) یا انفجار گرادیان ( gradient explosion ) ناشی از افزایش عمق در شبکه‌های عصبی را کاهش می‌دهد و به ما کمک می‌کند تا ساختارهای Convolutional Neural Networks را با انعطاف‌پذیری بیشتری بسازیم. در فصل سوم بیشتر با این معماری آشنا می‌شویم.

## فصل دوم

### ۲- پیشینه پژوهش

مراحل کلی کار در تشخیص حالت چهره پیش پردازش ، استخراج ویژگی و دسته بندی می‌باشد که در ادامه به کارها و متدها و روش هایی که قبلا در هر مرحله انجام شده می‌پردازیم همچنین در ادامه روش‌های کلی که در برخی مقالات مرتبط انجام گرفته است را شرح می‌دهیم. پیش پردازش یک فرایند است که می‌تواند برای بهبود عملکرد سیستم FER استفاده شود و میتواند قبل از فرآیند استخراج ویژگی انجام شود[4]. پیش پردازش تصویر شامل انواع مختلفی از فرایندها مانند وضوح و مقیاس بندی تصویر ، تنظیم کنتراست و فرایندهای افزایش اضافی

است [5] برای بهبود فریم های بیان [6] است. فرآیند استخراج ویژگی مرحله بعدی سیستم FER است. استخراج ویژگی یافتن و به تصویر کشیدن ویژگی های مثبت مورد توجه در یک تصویر برای پردازش بیشتر است. در پردازش تصویر ، استخراج ویژگی بینایی کامپیوتر مرحله مهمی است ، در حالی که حرکت از تصویر گرافیکی به تصویر ضمنی را مشخص می کند. سپس از این داده ها میتوان به عنوان ورودی طبقه بندی استفاده کرد. روشهای استخراج ویژگی به پنج نوع تقسیم می شوند مانند روش مبتنی بر ویژگی بافت، روش مبتنی بر لبه ، روش مبتنی بر ویژگی جهانی و محلی ، روش مبتنی بر ویژگی هندسی و روش مبتنی بر وصله فیلتر گابور با ویژگی اندازه اطلاعات مربوط به سازماندهی تصویر چهره را محدود می کند ویژگی فازی اطلاعات مربوط به شرح کامل ویژگیهای بزرگی [10]، [9]، [8]، [7]، [5] الگوی باینری محلی LBP نیز توصیف کننده بافت است و میتوان از آن برای استخراج ویژگی استفاده کرد به طور کلی ویژگیهای LBP با کد دودویی تولید می شود و می توان آن را با استفاده از آستانه بین پیکسل مرکزی و پیکسلهای محلی آن به دست آورد [11]، [12]. طبقه بندی آخرین مرحله سیستم FER است که در آن طبقه بندی کننده عبارتهایی مانند لبخند ، غم ، غافلگیری ، عصبانیت ، ترس ، انزجار و خنثی را طبقه بندی می کند از روش خط فاصله DLHD برای تشخیص عبارات استفاده میشود. [13] معیار فاصله اقلیدسی نیز برای اهداف طبقه بندی استفاده می شود که از ماتریس نمره نرمال و شباهت برای برآورد فاصله اقلیدسی استفاده می کند [10]. Kahua و همکاران [14] برای تشخیص حالات چهره از شبکه های عصبی کانولوشن استفاده کردند و برنده چالش شناخت احساسات در طبیعت ۲۰۱۳ EmotiW شدند یکی دیگر از مدلهای CNN و به دنبال آن یک ماشین بردار پشتیبانی خطی توسط تانگ و همکارانش برای تشخیص حالات صورت آموزش داده شد. [15] این برنده چالش تشخیص چهره FER2013 شد [16] کاهو و همکاران [14] CNN ها را برای استخراج ویژگیهای بصری همراه با ویژگی های صوتی در نمایش داده های چند حالته اعمال کرد. Nezami و همکاران [17] از یک مدل CNN برای تشخیص حالات صورت استفاده کرد ، جایی که نمایش آموخته شده در مدل شرح تصاویر استفاده می شود. در سال ۲۰۱۹ Shervin Minnie و همکاران [18] چندین کار با استفاده از مدلهای یادگیری عمیق، چارچوبی را برای تشخیص حالت چهره به انتها ارائه کردند علیرغم عملکرد بهتر این آثار ، هنوز به نظر می رسد که فضای خوبی برای پیشرفت وجود دارد. در این روش ، یک روش یادگیری عمیق را بر اساس شبکه تحریک توجهی پیشنهاد کردند ، که قادر بود بر قسمتهای مهم صورت تمرکز کند و نسبت به مدلهای قبلی در مجموعه داده های متعدد، از جمله ۲۰۱۳-CK FER، FERG و JAFFE ، به پیشرفت قابل توجهی دست پیدا کرد. از یک روش تجسم استفاده کردند که قادر بود مناطق مهم چهره را برای تشخیص احساسات مختلف اساس خروجی طبقه بندی کننده پیدا کند. از طریق نتایج تجربی ، نشان دادند که به نظر می رسد احساسات مختلف نسبت به قسمتهای مختلف صورت حساس بودند. در این کار از یک چارچوب مبتنی بر یادگیری عمیق را پیشنهاد کردند.

## فصل سوم

### ۳- روش پیشنهادی

#### ۱-۳ پیش پردازش

بر روی دیتاست موجود، با استفاده از تکنیک ImageDataGenerator فراهم شده در Keras، داده‌افزایی‌های زیر را بر روی دیتاست آموزش (train) اعمال کرده ایم.

جدول ۲- داده‌افزایی‌های انجام شده بر روی داده‌ها

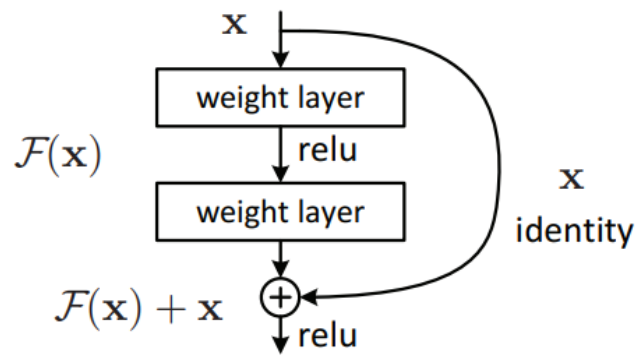
مقدار	نوع داده‌افزایی
۰,۱	انتقال در راستای عرض (range shift width)
۰,۱	انتقال در راستای ارتفاع (range shift height)
صحیح (True)	قرینه‌سازی افقی (flip horizontal)
۱/۲۵۵	تغییر اندازه (rescale)

با استفاده از تکنیک ImageDataGenerator، می‌توانیم داده‌افزایی را هم زمان با اجرا کردن مدل (Real Time augmentation) انجام دهیم که بسیار می‌تواند استفاده از حافظه (Memory) را نسبت به داده‌افزایی‌های سنتی کاهش دهد.

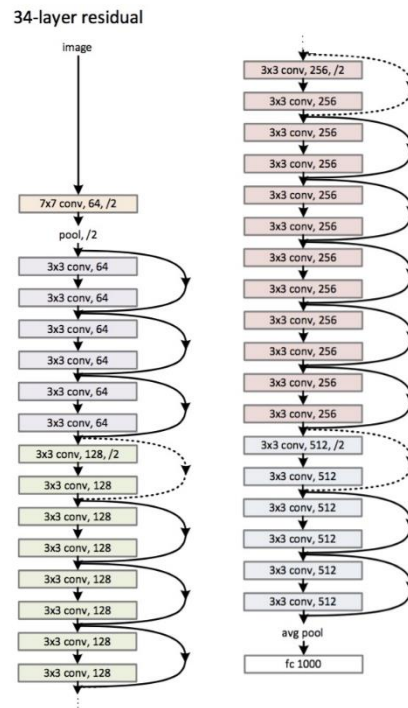
#### ۲-۳ مدل ابتدایی ResNet34

طی چند سال اخیر، در حوزه‌ی بینایی ماشین (Computer Vision) پیشرفت‌های چشمگیری صورت گرفته است؛ به‌خصوص با معرفی شبکه‌های عصبی عمیق کانولوشن (CNN)، نتایج بسیار خوبی را در زمینه‌ی مسائلی مانند طبقه‌بندی تصاویر (Image Classification) و شناسایی تصاویر (Image Recognition) به دست آورده‌ایم. طی سال‌ها محققان به ایجاد شبکه‌های عصبی عمیق‌تر (افزودن لایه‌های بیشتر) برای حل و بهبود چنین کارهای پیچیده‌ای تمایل پیدا کرده‌اند، اما موضوع این است که با افزودن لایه‌های بیشتری به شبکه‌ی عصبی، آموزش آن‌ها دشوار می‌شود و دقت عملکرد شبکه شروع به کاهش می‌کند.

اتصالات میانبر (Skip Connections) یا اتصالات اضافی (Residual Connections) راه‌حلی بود که شبکه رزنت (ResNet) برای حل مشکل شبکه‌های عمیق ارائه کرد. در شکل زیر یک بلاک اضافی (Residual Block) را مشاهده می‌کنیم. همان‌طور که مشخص است، فرق این شبکه با شبکه‌های معمولی این است که یک اتصال میانبر دارد که از یک یا چند لایه عبور می‌کند و آن‌ها را در نظر نمی‌گیرد؛ درواقع به‌نوعی میانبر می‌زند و یک لایه را به لایه‌ی دورتر متصل می‌کند.



شکل ۲- اتصال میانبر در معماری ResNet34



شکل 3 - معماری شبکه عصبی ResNet34

درفصل بعدی به پیاده سازی و ارزیابی این مدل بر روی دیتاست مذکور می پردازیم.

### ۳-۳ مدل پیشنهادی اول

این مدل بر پایه مدل ResNet34 طراحی شده است و از ویژگی های بلاک اضافی (Residual Block) برای غلبه بر مشکل محوشدگی گرادیان (Vanishing Gradient) یا انفجار گرادیان (Exploding Gradient) بهره برده است.

این مدل همچنین از تکنیک Dropout برای غلبه بر مشکل overfitting در چندین لایه بهره برده است. Dropout روشی برای منظم سازی (Regularization) است و خطای تعمیم پذیری (Generalization error) را با کاهش گنجایش مدل کمتر می کند؛ بدین ترتیب که در هر دور آموزشی، به جای استفاده از همه نورون ها، تنها برخی از نورون ها (با احتمال  $p$ ) فعال می شوند.

```
ZeroPadding2D((3, 3))(x_input) → Conv2D(64, kernel_size=7, strides=2, padding='same')
→ BatchNormalization() → Activation('relu')
→ MaxPool2D(pool_size=3, strides=2, padding='same') → identity_block(64)
→ identity_block(64) → identity_block(64) → convolutional_block(128) → Dropout( 0.25)
→ identity_block(128) → identity_block(128) → Dropout( 0.25) → identity_block(128)
→ identity_block(128) → identity_block(128) → convolutional_block(256) → Dropout( 0.25)
→ identity_block(256) → identity_block(256) → identity_block(256) → Dropout( 0.25)
→ identity_block(256) → identity_block(256) → identity_block(256) → Dropout( 0.25)
→ identity_block(256) → identity_block(256) → identity_block(256) → Dropout( 0.25)
→ identity_block(256) → identity_block(256) → convolutional_block(512) → Dropout( 0.25)
→ identity_block(512) → identity_block(512) → identity_block(512) → Dropout( 0.25)
→ identity_block(512) → identity_block(512) → identity_block(512) → Dropout( 0.25)
→ identity_block(512) → AveragePooling2D((2,2), padding = 'same') → Flatten() → Dense(512,
activation = 'relu') → Dense(classes, activation='softmax')
```

convolutional\_block و identity\_block مطابق با تعاریف مقاله ResNet34 می باشند.

### ۴-۳ مدل پیشنهادی دوم

طراحی شبکه عصبی، مشابه با مدل شماره یک است و تنها تفاوت در وزن دار کردن کلاس های دیتاست در محاسبه خطا نسبت به فراوانی نمونه های هر کلاس می باشد که به منظور بهبود عمل کرد در دسته بندی داده های کلاس های اقلیت انجام شده است.

جدول ۳- ضریب های وزنی اعمال شده در هر یک از کلاس ها در دسته بندی پیشنهادی دوم

ضریب وزنی در محاسبه خطا	تعداد داده	کلاس احساسی
۱,۰۲	۴۹۵۳	عصبانی (Anger)
۹,۴۰	۵۴۷	انزجار (Disgust)
۱,۰۰	۵۱۲۱	ترس (Fear)
۰,۵۶	۸۹۸۹	خوشحالی (Happiness)
۰,۸۲	۶۰۷۷	ناراحتی (Sadness)
۰,۸۴	۴۰۰۲	تعجب (Surprise)
۱,۲۹	۶۱۹۸	خنثی (Neutral)

### ۵-۳ EfficientNet

EfficientNet یک روش شبکه عصبی کانولوشنال و روش مقیاس گذاری است که از یک ضریب ترکیبی استفاده میکند تا عرض، عمق و وضوح شبکه را به صورت اصولی مقیاس بندی کند. روش مقیاس گذاری EfficientNet به طور یکنواخت عرض، عمق و وضوح شبکه را با مجموعه های از ضرایب ثابت مقیاس بندی میکند. به عنوان مثال، اگر بخواهیم منابع محاسباتی  $2^N$  برابر بیشتر استفاده کنیم، بنابراین میتوانیم عمق شبکه را با  $\alpha^N$  عرض توسط  $\beta^N$  و اندازه تصویر را توسط  $\gamma^N$  افزایش دهیم، که  $\beta, \gamma$  ،  $\alpha$  ضرایب ثابت هستند. روش مقیاس بندی مرکب با این بینش توجیه میشود اگر تصویر ورودی بزرگتر باشد، شبکه به لایه های بیشتری برای افزایش زمینه پذیرش و کانال های بیشتری برای جذب الگوهای ریز ساختار در تصویر بزرگتر نیاز دارد. پایه شبکه EfficientNet-B بر اساس بلوک های باقیمانده گلوگاه معکوس است.

در این پژوهش از مدل های EfficientNet-B3 و EfficientNet-B5 استفاده شده است [19].



## فصل چهارم

### ۴- پیاده‌سازی مدل‌ها و تحلیل و ارزیابی نتایج

#### ۱-۴ مقدمه

در این فصل به بررسی و مقایسه‌ی نتایج مدل‌های ارائه‌شده در فصل قبل می‌پردازیم.

#### ۲-۴ مدل ابتدایی ResNet34

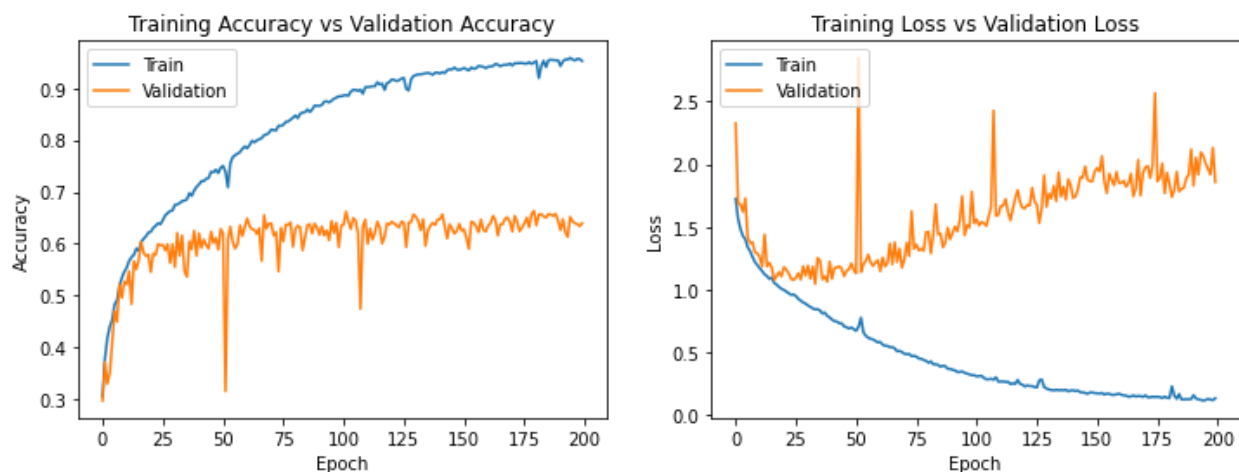
مدل ذکر شده را مطابق با معماری طراحی شده در مقاله اصلی پیاده‌سازی و با مشخصات زیر تنظیم (compile) کرده ایم.

بهینه‌ساز (optimizer): Adam

تابع خطا (loss): categorical\_crossentropy

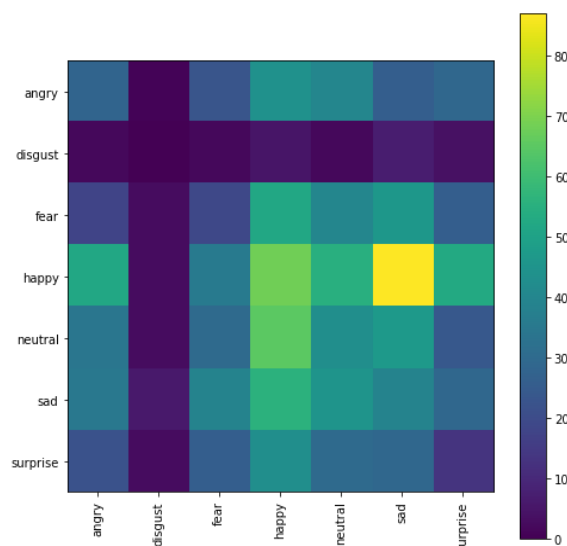
سپس به تعداد دوره (epoch) ۲۰۰ و اندازه‌ی بسته ۶۴، بر روی دیتاست FER-2013 آموزش (train) داده ایم که دقت‌های زیر را بر روی دیتاست آموزشی و آزمایشی به دست آورده است.

final train accuracy = 94.36 , validation accuracy = 63.97

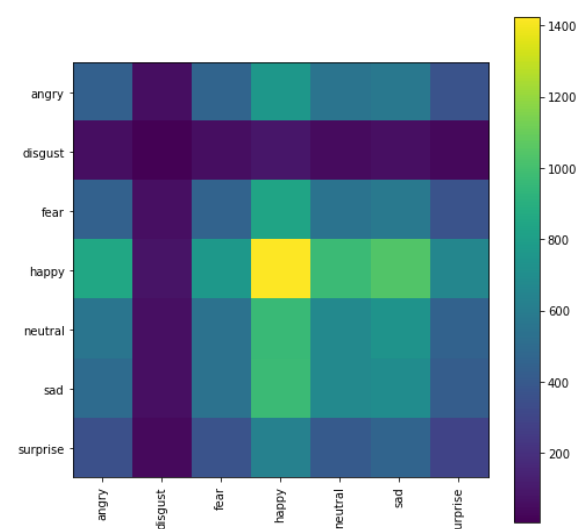


شکل ۴- نمودار تغییرات خطا در دوره‌ها (سمت راست) و نمودار تغییرات دقت در دوره‌ها (سمت چپ)

همچنین ماتریس های درهم ریختگی زیر را برای دیتاست آموزشی و آزمایشی حاصل می دهد.

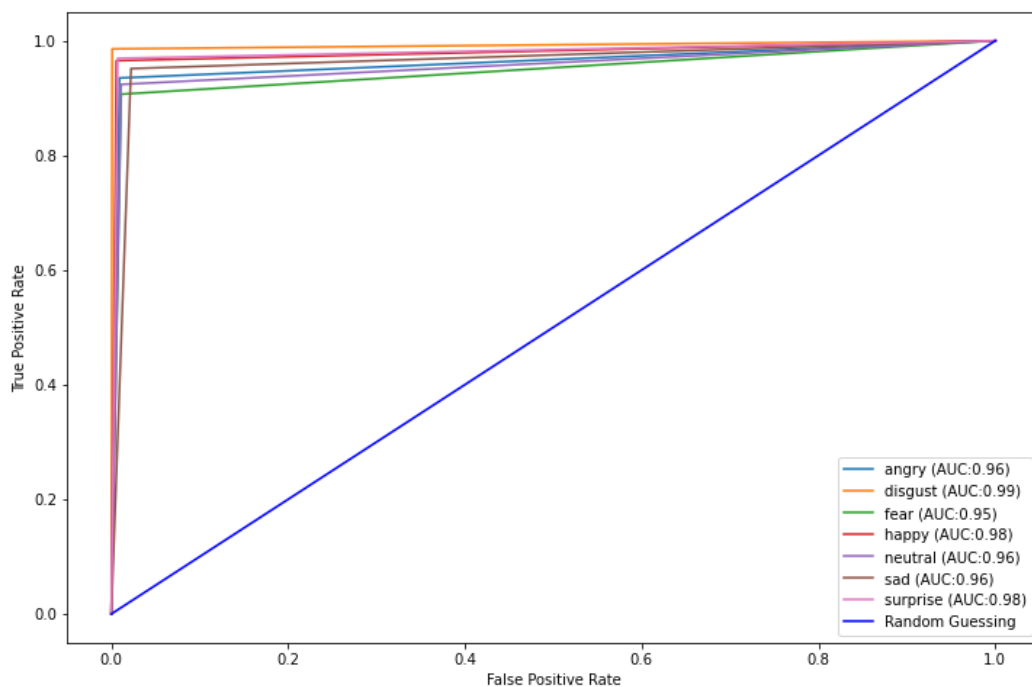


شکل ۵- ماتریس درهم ریختگی ResNet34 بر روی داده های آزمایشی (validation)



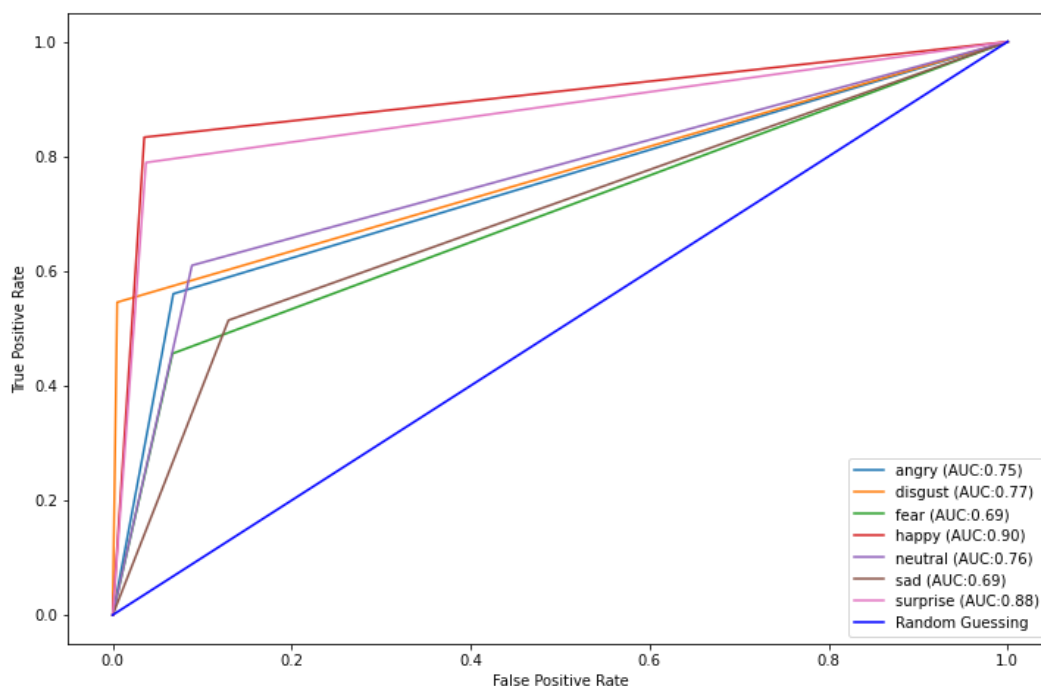
شکل ۴- ماتریس درهم ریختگی ResNet34 بر روی داده های آموزشی (Train)

نمودار ROC برای داده های آموزشی (Train) به شکل زیر می باشد.



شکل ۶- نمودار ROC مدل ResNet34 بر روی داده های آموزشی

نمودار ROC برای داده‌های آزمایشی (Test) به شکل زیر می‌باشد.



شکل ۷- نمودار ROC مدل ResNet34 بر روی داده‌های آزمایشی

#### ۳-۴ مدل پیشنهادی اول

مدل ذکر شده را مطابق با طراحی بیان شده در فصل قبل پیاده سازی و با مشخصات زیر تنظیم (compile) کرده ایم.

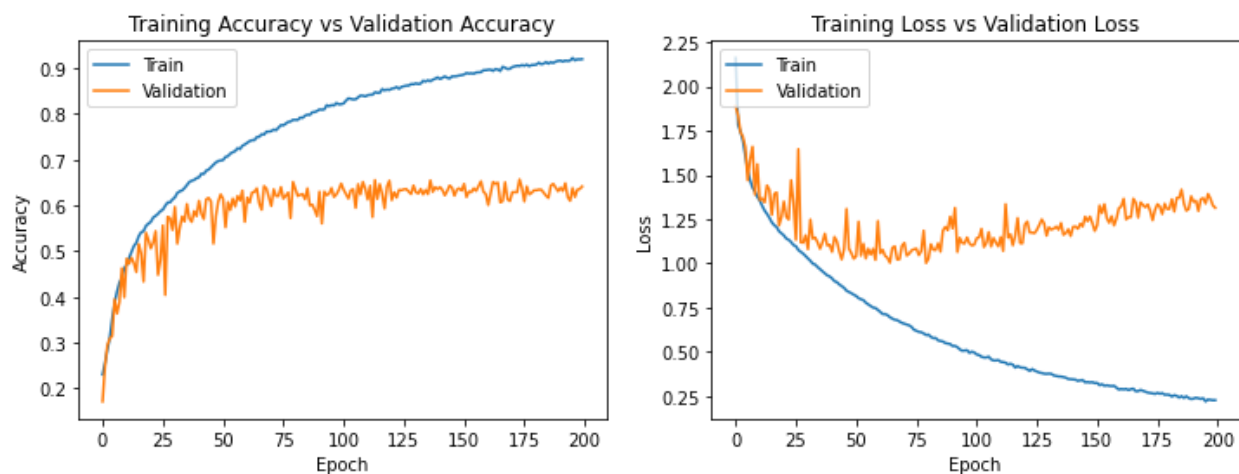
بهینه ساز (optimizer): Adam

تابع خطا (loss): categorical\_crossentropy

همچنین لازم به ذکر می‌باشد که برای این مدل از ModelCheckpoint برای نگهداری بهترین validation accuracy در طول epochها استفاده شده است.

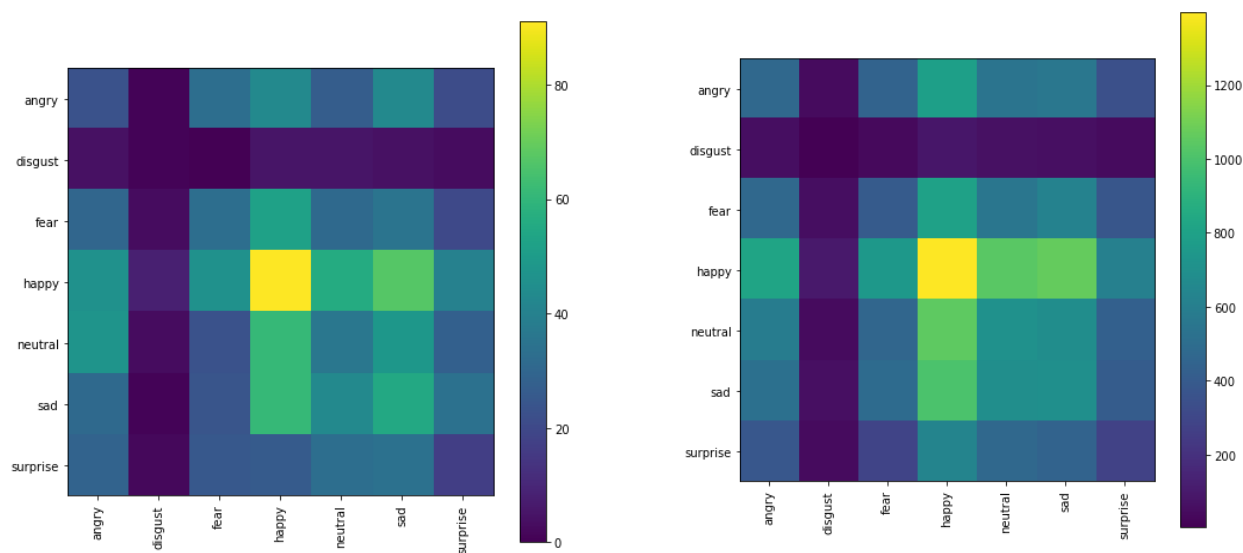
سپس به تعداد دوره (epoch) ۲۰۰ و اندازه‌ی بسته ۶۴، بر روی دیتاست FER-2013 آموزش (train) داده ایم که دقت های زیر را بر روی دیتاست آموزشی و آزمایشی به دست آورده است.

final train accuracy = 92.23 , validation accuracy = 65.71



شکل ۹- نمودار تغییرات خطا در دوره‌ها (سمت راست) و نمودار تغییرات دقت در دوره‌ها (سمت چپ)

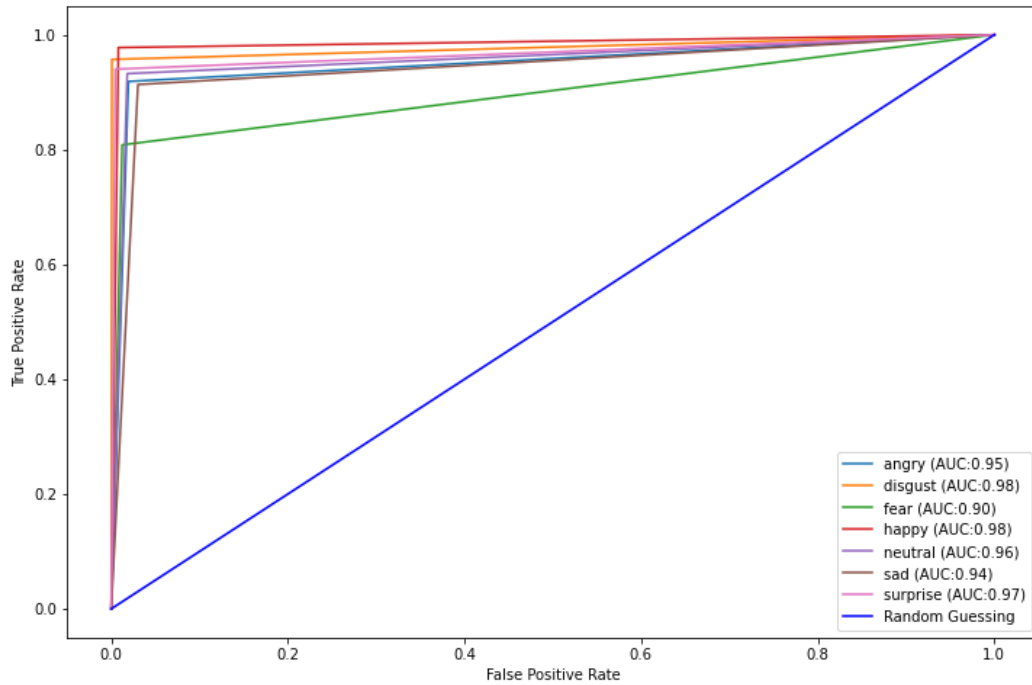
همچنین ماتریس های درهم ریختگی زیر را برای دیتاست آموزشی و آزمایشی حاصل می‌دهد.



شکل ۸- ماتریس درهم‌ریختگی مدل پیشنهادی اول بر روی داده‌های آزمایشی (Validation)

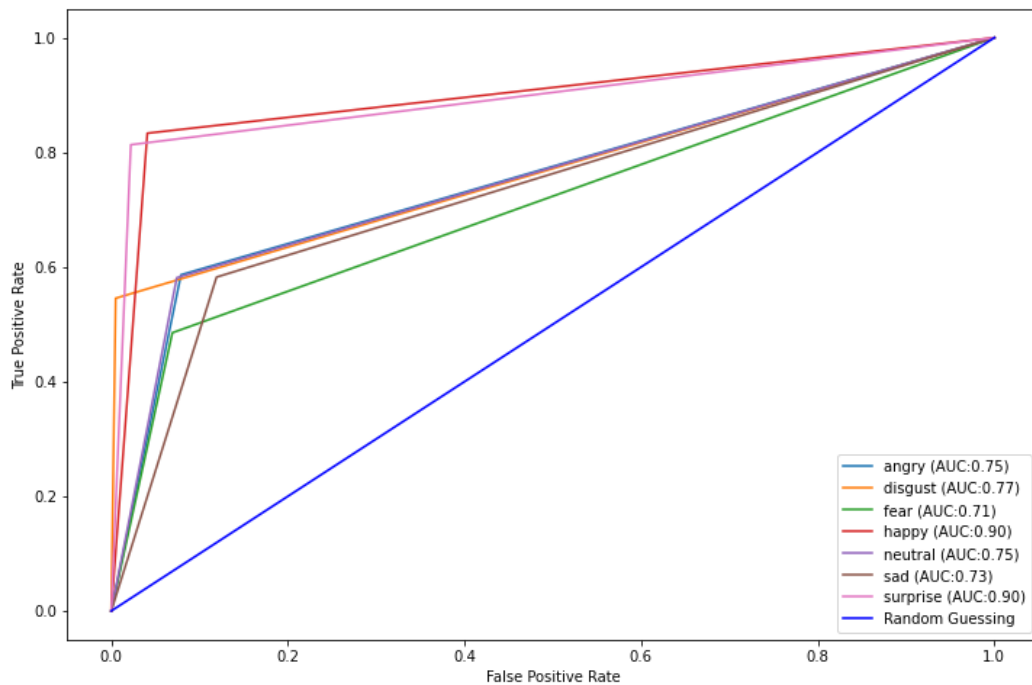
شکل ۹- ماتریس درهم‌ریختگی مدل پیشنهادی اول بر روی داده‌های آموزشی (Train)

نمودار ROC برای داده‌های آموزشی (Train) به شکل زیر می‌باشد.



شکل ۱۰- نمودار ROC مدل پیشنهادی اول بر روی داده‌های آموزشی (Train)

نمودار ROC برای داده‌های آزمایشی (Test) به شکل زیر می‌باشد.



شکل ۱۱- نمودار ROC مدل پیشنهادی اول بر روی داده‌های آزمایشی (Validation)

#### ۴-۴ مدل پیشنهادی دوم

مدل ذکر شده را مطابق با طراحی بیان شده در فصل قبل پیاده سازی و با مشخصات زیر تنظیم (compile) کرده ایم.

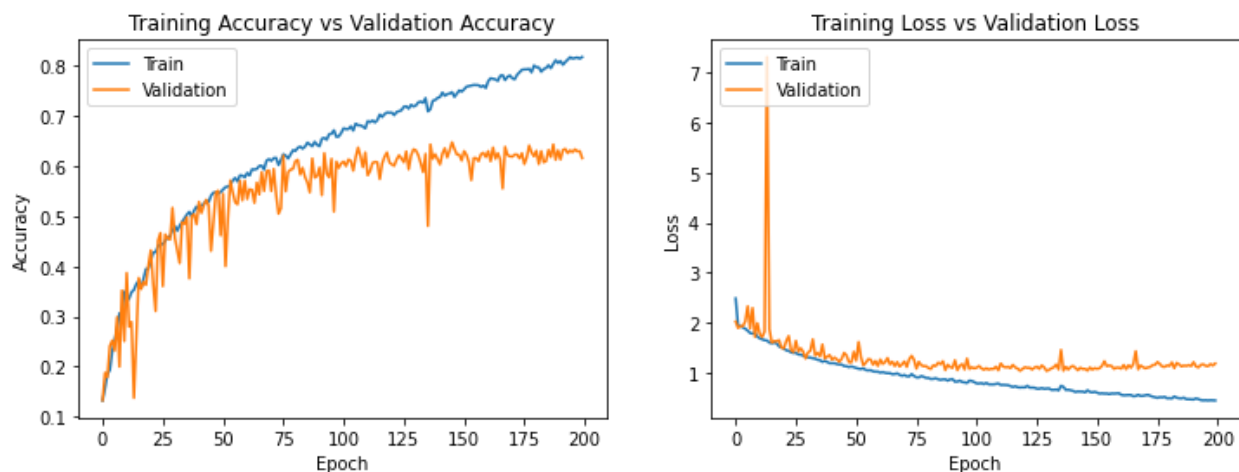
بهینه ساز (optimizer): Adam

تابع خطا (loss): categorical\_crossentropy

همچنین لازم به ذکر می باشد که برای این مدل از ModelCheckpoint برای نگهداری بهترین validation accuracy در طول epoch ها استفاده شده است.

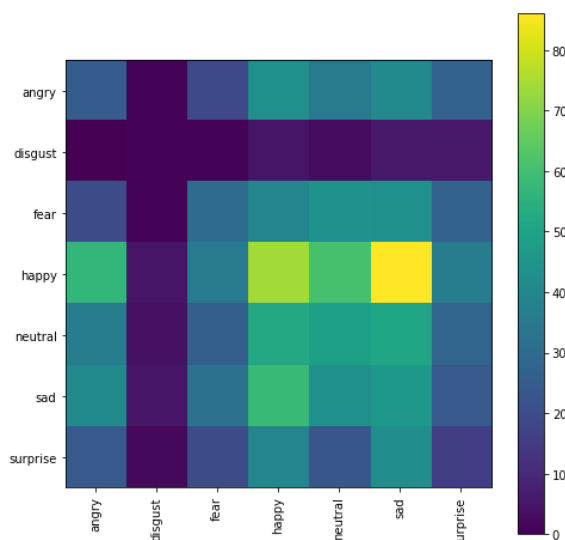
سپس به تعداد دوره (epoch) ۲۰۰ و اندازه ی بسته ۶۴، بر روی دیتاست FER-2013 آموزش (train) داده ایم که دقت های زیر را بر روی دیتاست آموزشی و آزمایشی به دست آورده است.

final train accuracy = 75.61 , validation accuracy = 64.73

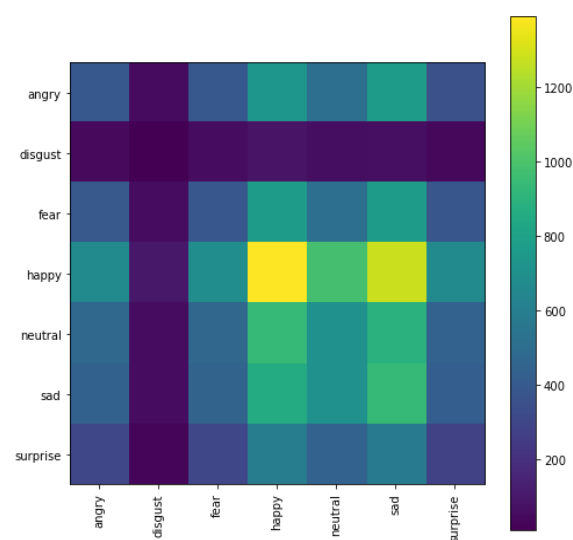


شکل ۱۲- نمودار تغییرات دقت در دوره ها (سمت چپ) و نمودار تغییرات خطا در دوره ها (سمت راست)

همچنین ماتریس های درهم ریختگی زیر را برای دیتاست آموزشی و آزمایشی حاصل می دهد.

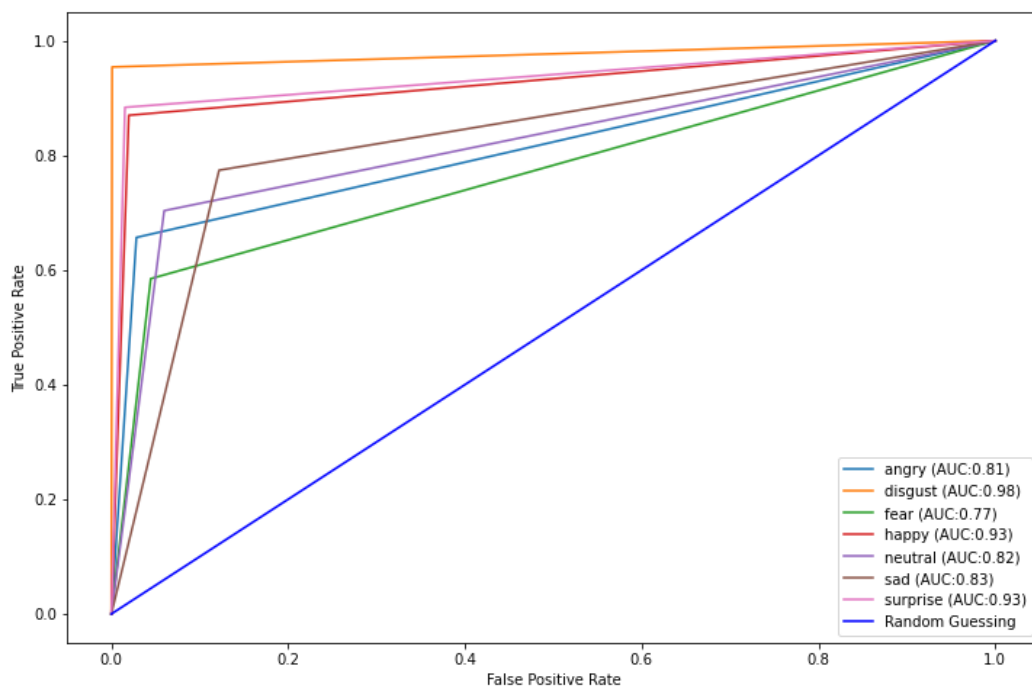


شکل ۱۲- ماتریس درهم ریختگی مدل پیشنهادی دوم بر روی داده های آزمایشی (Validation)



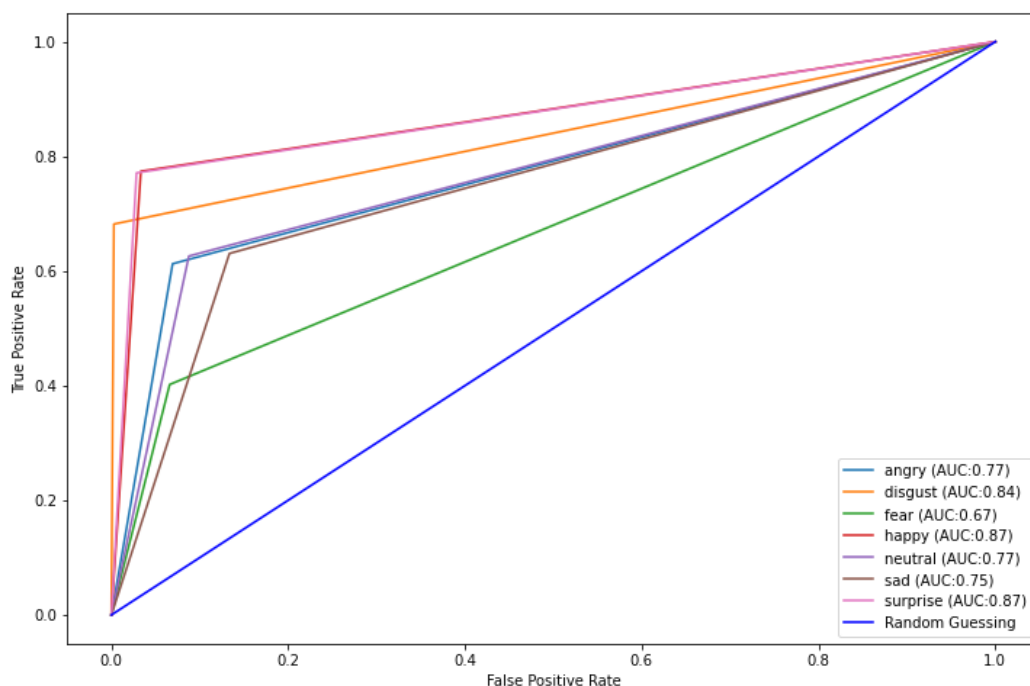
شکل ۱۳- ماتریس درهم ریختگی مدل پیشنهادی دوم بر روی داده های آموزشی (Train)

نمودار ROC برای داده های آموزشی (Train) به شکل زیر می باشد.



شکل ۱۴- نمودار ROC مدل پیشنهادی دوم بر روی داده های آموزشی (Train)

نمودار ROC برای داده‌های آزمایشی (Test) به شکل زیر می‌باشد.



شکل ۱۵- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آزمایشی (Validation)

#### ۵-۴ EfficientNet-B3

مدل EfficientNet-B3 را با استفاده از مدل‌های از پیش طراحی شده ماژول applications در keras با مشخصات زیر تنظیم (compile) کرده ایم.

بهینه ساز (optimizer): Adam

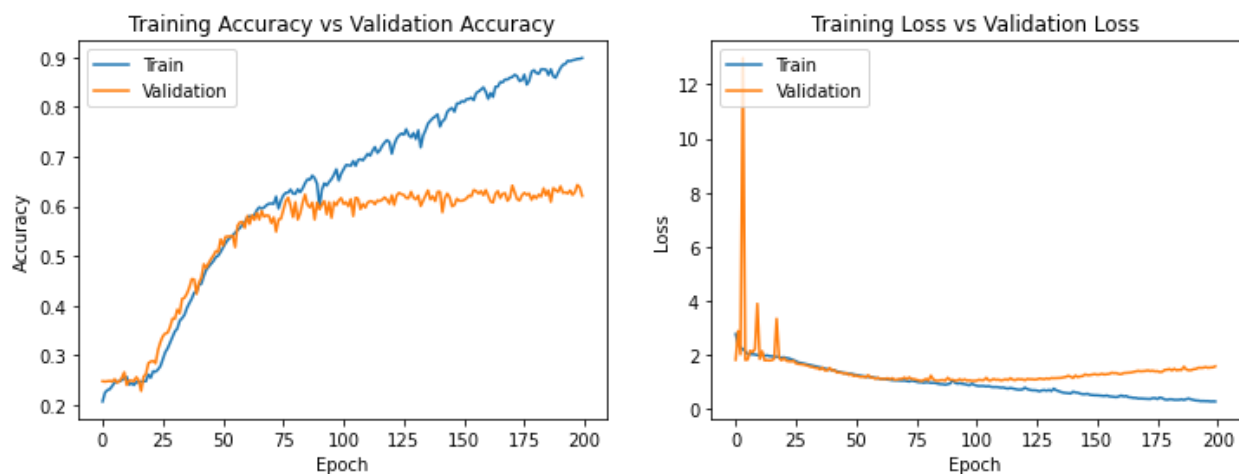
تابع خطا (loss): categorical\_crossentropy

همچنین لازم به ذکر می‌باشد که برای این مدل از ModelCheckpoint برای نگهداری بهترین validation accuracy در طول epochها استفاده شده است.

سپس به تعداد دوره (epoch) ۲۰۰ و اندازه‌ی بسته ۶۴، بر روی دیتاست FER-2013 آموزش (train) داده ایم که دقت‌های زیر را بر روی دیتاست آموزشی و آزمایشی به دست آورده است.

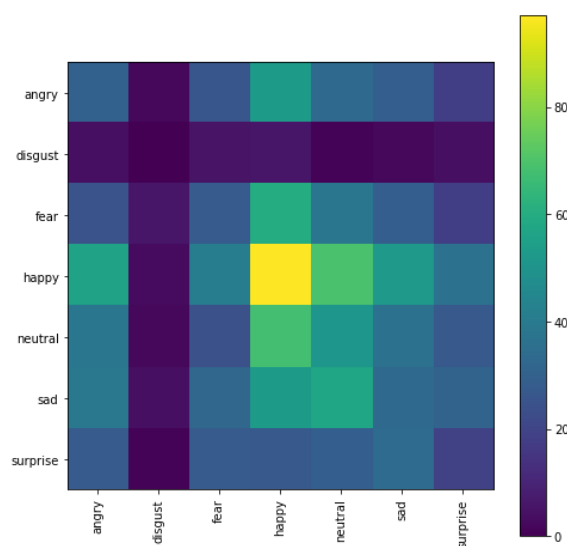


final train accuracy = 94.61 , validation accuracy = 64.32

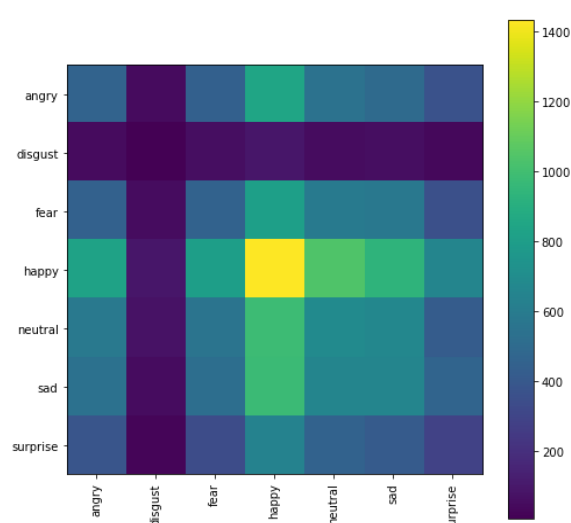


شکل ۱۶- نمودار تغییرات خطا در دوره‌ها (سمت راست) و نمودار تغییرات دقت در دوره‌ها (سمت چپ)

همچنین ماتریس‌های درهم ریختگی زیر را برای دیتاست آموزشی و آزمایشی حاصل می‌دهد.

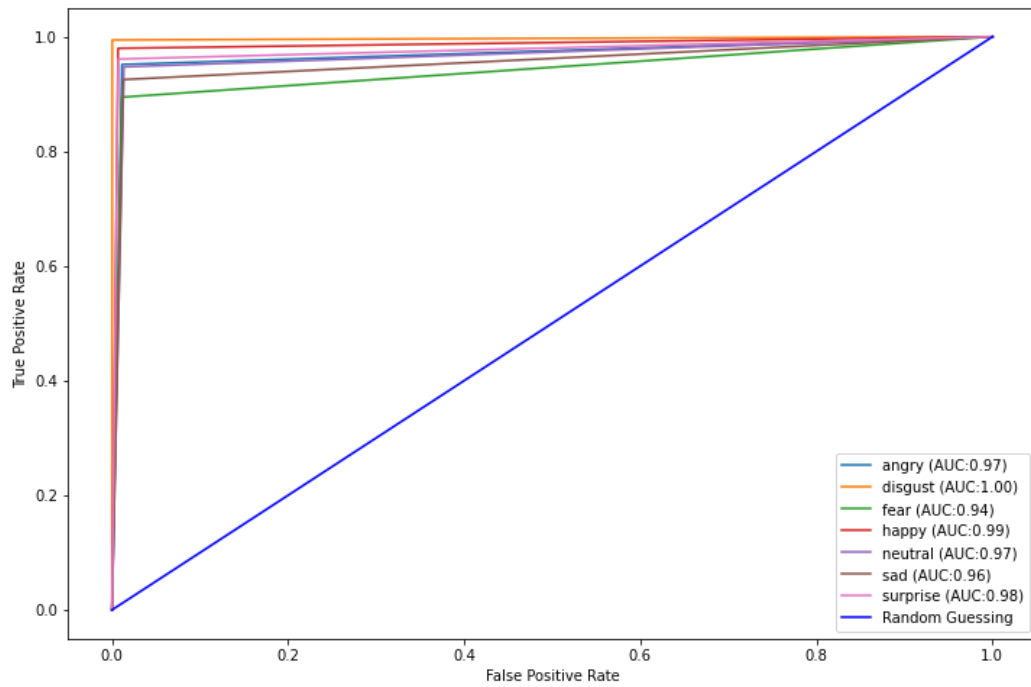


شکل ۱۶- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آزمایشی (Validation)



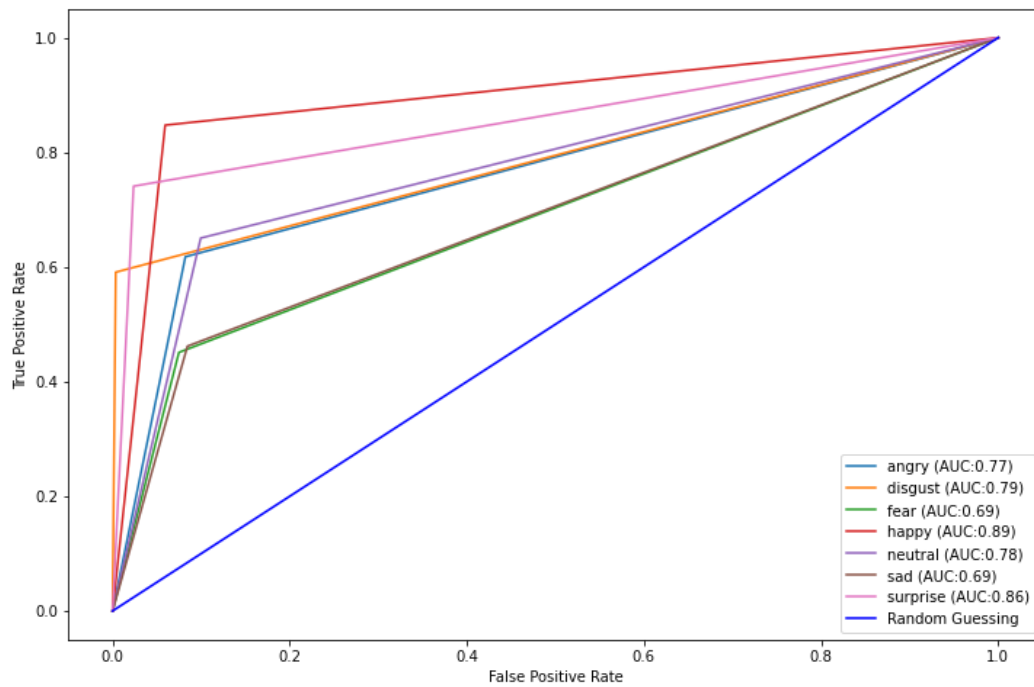
شکل ۱۷- ماتریس درهم‌ریختگی مدل پیشنهادی دوم بر روی داده‌های آموزشی (Train)

نمودار ROC برای داده‌های آموزشی (Train) به شکل زیر می‌باشد.



شکل ۱۸- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آموزشی (Train)

نمودار ROC برای داده‌های آزمایشی (Test) به شکل زیر می‌باشد.



شکل ۱۹- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آزمایشی (Validation)

#### ۶-۴ EfficientNet-B5

مدل EfficientNet-B5 را با استفاده از مدل های از پیش طراحی شده ماژول applications در keras با مشخصات زیر تنظیم (compile) کرده ایم.

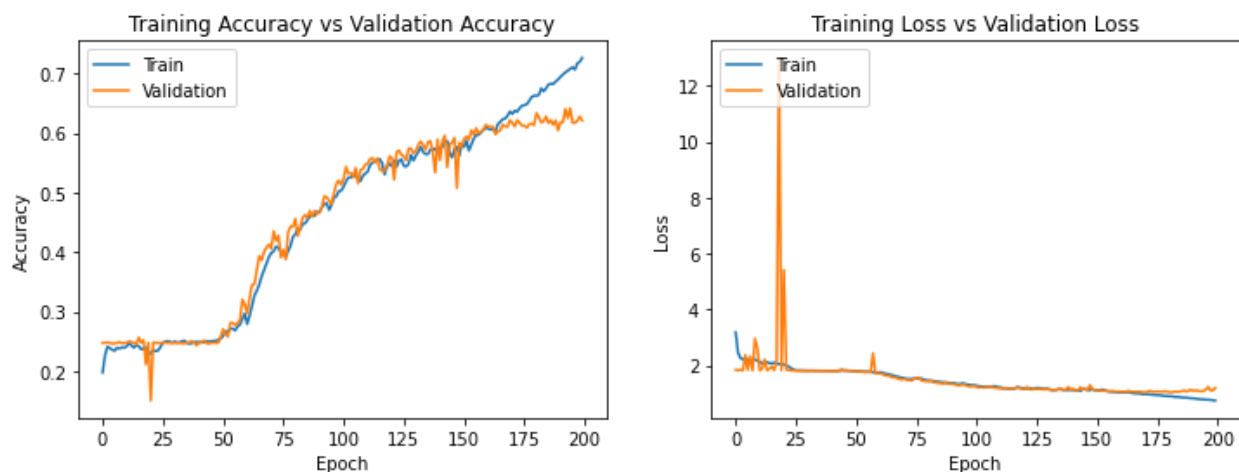
بهینه ساز (optimizer): Adam

تابع خطا (loss): categorical\_crossentropy

همچنین لازم به ذکر می باشد که برای این مدل از ModelCheckpoint برای نگهداری بهترین validation accuracy در طول epoch ها استفاده شده است.

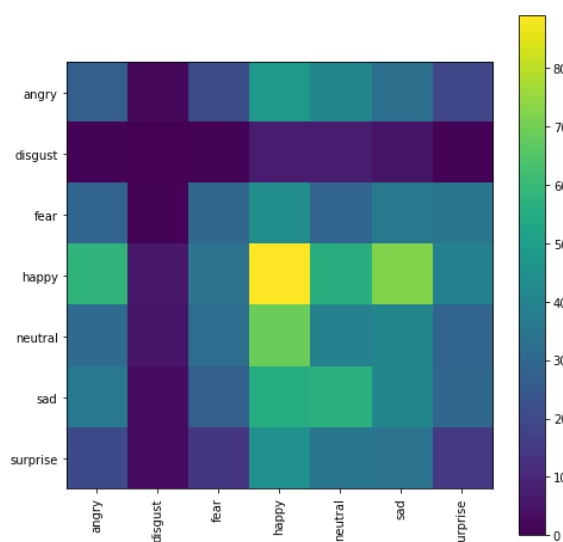
سپس به تعداد دوره (epoch) ۲۰۰ و اندازه ی بسته ۶۴، بر روی دیتاست FER-2013 آموزش (train) داده ایم که دقت های زیر را بر روی دیتاست آموزشی و آزمایشی به دست آورده است.

final train accuracy = 74.42 , validation accuracy = 64.18

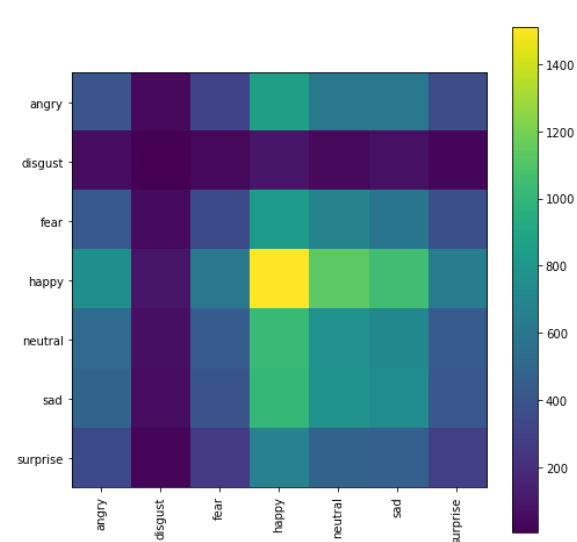


شکل ۲۰- نمودار تغییرات دقت در دوره ها (سمت راست) و نمودار تغییرات دقت در دوره ها (سمت چپ)

همچنین ماتریس های درهم ریختگی زیر را برای دیتاست آموزشی و آزمایشی حاصل می دهد.

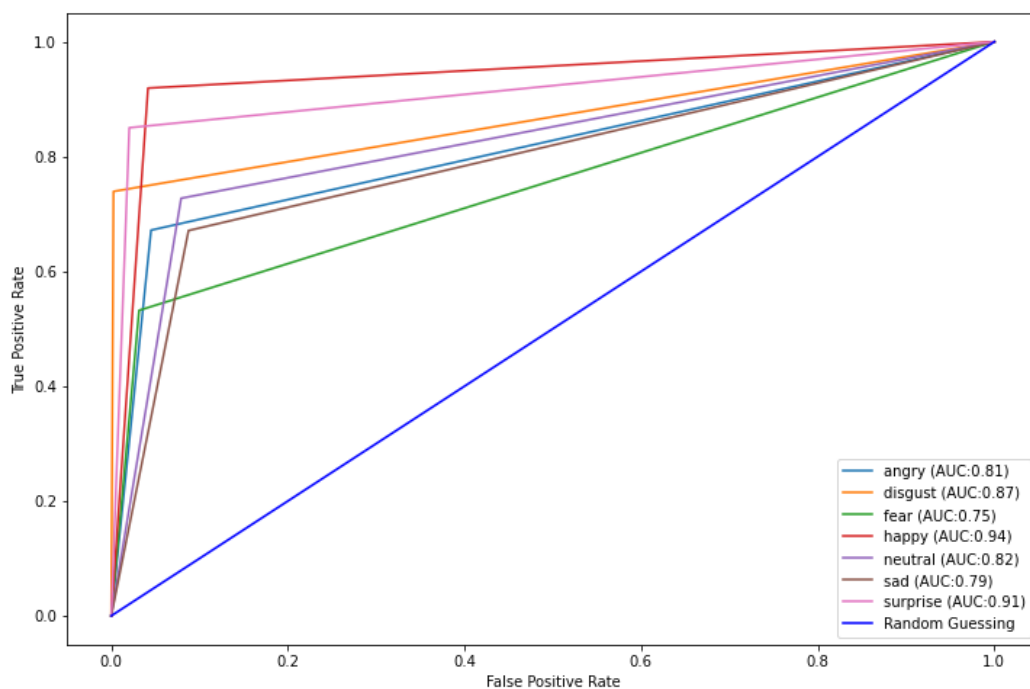


شکل ۲۲- ماتریس درهم ریختگی مدل پیشنهادی دوم بر روی داده های آزمایشی (Validation)



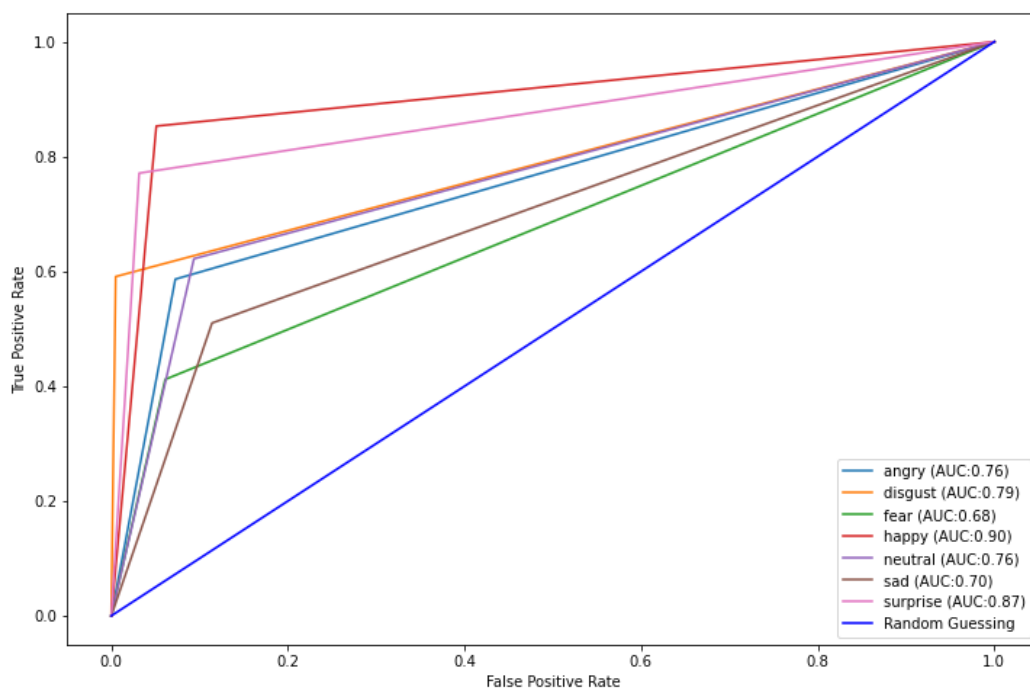
شکل ۲۱- ماتریس درهم ریختگی مدل پیشنهادی دوم بر روی داده های آزمایشی (Validation)

نمودار ROC برای داده های آموزشی (Train) به شکل زیر می باشد.



شکل ۲۳- نمودار ROC مدل پیشنهادی دوم بر روی داده های آموزشی (Train)

نمودار ROC برای داده‌های آزمایشی (Test) به شکل زیر می‌باشد.



شکل ۲۴- نمودار ROC مدل پیشنهادی دوم بر روی داده‌های آموزشی (Train)

#### ۷-۴ ارزیابی مدل‌ها

در جدول زیر خلاصه‌ای از نتایج مدل‌های مطرح شده بر روی دیتاست آزمایشی (validation set) و دیتاست آموزشی (Train set) آورده شده است.

جدول ۴- دقت مدل‌های مطرح شده بر روی داده‌های آموزشی (Train) و آزمایشی (Validation)

دسته‌بند	دقت بر روی دیتاست آموزشی (Train)	دقت بر روی دیتاست آزمایشی (Validation)
ResNet34	94.36	63.97
پیشنهادی اول	92.23	65.71
پیشنهادی دوم	75.61	64.73
EfficientNet-B3	94.61	64.32
EfficientNet-B5	74.42	64.18

با توجه به جدول قبل، هردو دسته‌بند پیشنهادی توانسته‌اند دقت بالاتری در دیتاست آزمایشی نسبت به مدل های ResNet34 و EfficientNet-B3 و EfficientNet-b5 به دست آورند. بیشترین دقت را مدل پیشنهادی اول به دلیل عدم توجه به کلاس های نامتوازن کسب کرده است در صورتی که اعمال کردن وزن های جدید به کلاس ها در مدل پیشنهادی دوم باعث شده‌است تا دقت این مدل به نسبت پایین‌تر از مدل پیشنهادی اول باشد.

همچنین در جدول زیر میزان معیار AUC برای هر کلاس در بر روی دیتاست آزمایشی (Validation) ذکر شده‌است.

جدول ۵- AUC مدل های مطرح شده بر روی کلاس‌های دیتاست FER-2013

	عصبانیت	انزجار	ترس	خوشحالی	ناراحتی	تعجب	خنثی
ResNet34	۰,۷۵	۰,۷۷	۰,۶۹	۰,۹۰	۰,۶۹	۰,۸۸	۰,۷۶
پیشنهادی اول	۰,۷۵	۰,۷۷	۰,۷۱	۰,۹۰	۰,۷۳	۰,۹۰	۰,۷۵
پیشنهادی دوم	۰,۷۷	۰,۸۴	۰,۶۷	۰,۸۷	۰,۷۵	۰,۸۷	۰,۷۷
EfficientNet-B3	۰,۷۷	۰,۷۹	۰,۶۹	۰,۸۹	۰,۶۹	۰,۸۶	۰,۷۸
EfficientNet-B5	۰,۷۶	۰,۷۹	۰,۶۸	۰,۹۰	۰,۷۰	۰,۸۷	۰,۷۶

همانطور که در جدول بالا نمایش داده شده است، مدل پیشنهادی اول در تمام کلاس ها از مدل ResNet34 بهتر عمل کرده است و مدل پیشنهادی دوم در کلاس های اقلیت توانسته است AUC بهتری نسبت به مدل پیشنهادی اول با وجود دقت پایین‌تر این مدل نسبت به مدل پیشنهادی اول حاصل دهد.

## فصل پنجم

### ۵- نتیجه‌گیری و پیشنهادها

#### ۱-۵ مقدمه

در این فصل به بررسی نتایج حاصل از پژوهش می‌پردازیم و در پایان، پیشنهاداتی برای پژوهش‌های آتی ارائه می‌شود.

#### ۲-۵ بررسی نتایج پژوهش

در این پژوهش، توانستیم با عمیق‌تر کردن معماری شبکه عصبی Resnet34 و اضافه کردن لایه‌های dropout نتایج حاصل از این مدل را بر روی دیتاست FER-2013 بهبود دهیم و دقت انسان در این دیتاست را کسب کنیم. اضافه کردن لایه‌های dropout توانست تا حد بالایی به مدل ما در غلبه کردن بر مشکل overfitting معماری Resnet34 کمک کند و دقت بالاتری در تمام کلاس‌ها فراهم کند.

همچنین این مدل بهبود یافته با تکنیک وزن‌دار کردن کلاس‌ها به نفع کلاس‌های اقلیت توانست دقت بالاتری در این کلاس‌های اقلیت فراهم کند و تا حد خوبی به مشکل کلاس‌های بسیار نامتوازن دیتاست FER-2013 کمک کند و در این امر ثمر بخش واقع شود.

#### ۳-۵ پیشنهادها جهت مطالعات آتی

با توجه به موفقیت dropout در بهبود عملکرد شبکه عصبی، می‌توان از دیگر روش‌های regularization همچون L2 و L1 استفاده کرد و نتایج را بررسی کرد. همچنین می‌توان از روش‌های Up sampling و down sampling برای بهبود توازن کلاس‌ها استفاده نمود.

## References

- [1] Picard RW, Picard R. Affective computing (Vol. 252). MIT press Cambridge. "EEG-detected olfactory imagery to reveal covert consciousness in minimally conscious state." Brain injury. 1997;29(13-14):1729-35.
- [2] Barman A, Dutta P. Facial expression recognition using distance and shape signature features. Pattern Recognition Letters. 2021 May 1;145:254-61.
- [3] Goodfellow, I.J., et al. Challenges in Representation Learning: A report on three machine learning contests. arXiv e-prints, 2013.
- [4] Poursaberi A, Noubari HA, Gavrilova M, Yanushkevich SN. Gauss–Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. EURASIP Journal on Image and Video Processing. 2012 Dec;2012(1):1-3.
- [5] Revina IM, Emmanuel WS. A survey on human face expression recognition techniques. Journal of King Saud University-Computer and Information Sciences. 2021 Jul 1;33(6):619-28.
- [6] Siddiqi MH, Ali R, Sattar A, Khan AM, Lee S. Depth camera-based facial expression recognition system using multilayer scheme. IETE Technical Review. 2014 Jul 4;31(4):277-86.
- [7] Owusu E, Zhan Y, Mao QR. An SVM-AdaBoost facial expression recognition system. Applied intelligence. 2014 Apr 1;40(3):536-45.
- [8] Zhang L, Tjondronegoro D, Chandran V. Random Gabor based templates for facial expression recognition in images with facial occlusion. Neurocomputing. 2014 Dec 5;145:451-64.
- [9] Hernandez-Matamoros A, Bonarini A, EscamillaHernandez E, Nakano-Miyatake M, Perez-Meana H. A facial expression recognition with automatic segmentation of face regions. InInternational Conference on Intelligent Software Methodologies, Tools, and Techniques 2015 Sep 15 (pp. 529-540). Springer, Cham.
- [10] Hegde GP, Seetha M, Hegde N. Kernel locality preserving symmetrical weighted fisher discriminant analysis based subspace approach for expression recognition. Engineering science and technology, an international journal. 2016 Sep 1;19(3):1321-33.



- [11] Cossetin MJ, Nievola JC, Koerich AL. Facial expression recognition using a pairwise feature selection and classification approach. In 2016 International Joint Conference on Neural Networks (IJCNN) 2016 Jul 24 (pp. 5149-5155). IEEE.
- [12] Happy SL, Routray A. Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing. 2014 Dec 29;6(1):1-2.
- [13] Gao Y, Leung MK, Hui SC, Tananda MW. Facial expression recognition from line-based caricatures. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans. 2003 Sep 4;33(3):407-12.
- [14] Kahou SE, Pal C, Bouthillier X, Froumenty P, Gülçehre Ç, Memisevic R, Vincent P, Courville A, Bengio Y, Ferrari RC, Mirza M. Combining modality specific deep neural networks for emotion recognition in video. In Proceedings of the 15th ACM on International conference on multimodal interaction 2013 Dec 9 (pp. 543-550).
- [15] Tang Y. Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239. 2013 Jun 2.
- [16] Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y, Thaler D, Lee DH, Zhou Y. Challenges in representation learning: A report on three machine learning contests. In International conference on neural information processing 2013 Nov 3 (pp. 117-124). Springer, Berlin, Heidelberg
- [17] Nezami OM, Dras M, Anderson P, Hamey L. Face-cap: Image captioning using facial expression analysis. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2018 Sep 10 (pp. 226-240). Springer, Cham
- [18] Minaee S, Minaei M, Abdolrashidi A. Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors. 2021 Jan;21(9):3046
- [19] T. Mingxing, V. LeQuoc, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019