**عنوان**

**خرداد 1403**

فهرست مطالب

[مقدمه 1](#_Toc167726029)

[داده‌ها 1](#_Toc167726030)

[پیش‌پردازش 3](#_Toc167726031)

[مدل‌سازی با ResNet50V2 3](#_Toc167726032)

[نتایج 5](#_Toc167726033)

فهرست اشکال

[شکل ‏2‑1 تعداد داده‌های هر کلاس در الف- مجموعه‌ی آموزش ب- مجموعه‌ی تست 2](#_Toc167726071)

[شکل ‏2‑2 نمونه‌ای از تصاویر مربوط به هر کلاس 2](#_Toc167726072)

[شکل ‏4‑1 معماری شبکه‌ی ResNet-50 v2 4](#_Toc167726073)

[شکل ‏5‑1 نمودار مربوط به loss و Accuracy 6](#_Toc167726074)

[شکل ‏5‑2 Confusion Matrix مربوط به خروجی 6](#_Toc167726075)

[شکل ‏5‑3 لیبل تشخیص داده شده توسط شبکه برای تصاویر مختلف به همراه لیبل واقعی 7](#_Toc167726076)

# مقدمه

تشخیص احساسات از روی چهره یکی از مسائل مهم و چالش‌برانگیز در حوزه‌ی پردازش تصویر و بینایی ماشین است. انسان‌ها به طور طبیعی قادر به تشخیص احساسات مختلف از روی چهره‌ی دیگران هستند، اما برای سیستم‌های کامپیوتری این فرآیند بسیار پیچیده‌تر است. با پیشرفت تکنولوژی و ظهور یادگیری عمیق، امکان پیاده‌سازی سیستم‌هایی که می‌توانند احساسات را با دقت بالا تشخیص دهند، فراهم شده است.

در این پروژه، هدف استفاده از تکنیک‌های یادگیری عمیق برای شناسایی احساسات مختلف از روی تصاویر چهره است. این سیستم‌ها می‌توانند در کاربردهای مختلفی مانند تشخیص حالات روحی در روان‌شناسی، بهبود تعامل انسان و کامپیوتر، سیستم‌های امنیتی و نظارتی و حتی در تحلیل بازخورد مشتریان در کسب‌وکارها مفید باشند.

# داده‌ها

دیتاست FER-2013 یکی از معتبرترین و پرکاربردترین دیتاست‌ها در زمینه‌ی تشخیص احساسات چهره است که در این پروژه مورد استفاده قرار گرفته است. این دیتاست در مسابقه‌ی "The International Conference on Machine Learning (ICML)" سال 2013 معرفی شده و شامل 35,887 تصویر از چهره‌ی افراد با ابعاد 48\*48 پیکسل است. تصاویر به صورت خاکستری بوده و هر تصویر به یکی از هفت کلاس احساسی زیر دسته‌بندی شده است:

1. خشم (Angry)

2. انزجار (Disgust)

3. ترس (Fear)

4. خوشحالی (Happy)

5. غم (Sad)

6. تعجب (Surprise)

7. بی‌تفاوتی (Neutral)

این دیتاست به سه مجموعه تقسیم شده است:

- مجموعه‌ی آموزشی (Training Set): شامل 28,709 تصویر برای آموزش مدل.

- مجموعه‌ی اعتبارسنجی (Validation Set): شامل 3,589 تصویر برای تنظیم هایپرپارامترها و جلوگیری از بیش‌برازش.

- مجموعه‌ی آزمون (Test Set): شامل 3,589 تصویر برای ارزیابی نهایی عملکرد مدل.

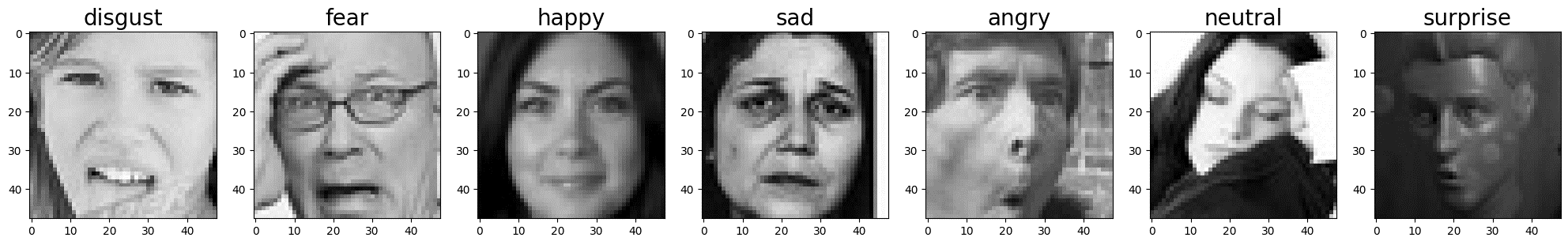
استفاده از FER-2013 به دلیل تعداد زیاد تصاویر و تنوع حالات چهره، این امکان را فراهم می‌آورد که مدل‌های یادگیری عمیق بتوانند ویژگی‌های پیچیده و دقیق را از تصاویر استخراج کرده و با دقت بالایی احساسات را تشخیص دهند.

در شکل ‏2‑1 تعداد داده‌های مرتبط با هر کلاس برای مجموعه‌ی تست و آموزش نشان داده شده است.

|  |  |
| --- | --- |
| الف | ب |

شکل ‏2‑1 تعداد داده‌های هر کلاس در الف- مجموعه‌ی آموزش ب- مجموعه‌ی تست

در شکل ‏2‑2 نیز یک نمونه از تصاویر هر کلاس نشان داده شده است.



شکل ‏2‑2 نمونه‌ای از تصاویر مربوط به هر کلاس

برای دسترسی به دیتاست مورد نظر بر روی [لینک](https://www.kaggle.com/datasets/msambare/fer2013) کلیک کنید.

# پیش‌پردازش

در این پروژه، از دو فرآیند پیش‌پردازش برای داده‌های آموزشی و داده‌های آزمون استفاده شد:

**1. پیش‌پردازش داده‌های آموزشی:**

- نرمال‌سازی: تصاویر با مقیاس 255/1 نرمال‌سازی شدند تا پیکسل‌ها در محدوده [0, 1] قرار گیرند.

- افزایش داده‌ها: شامل چرخش تصویر به میزان 10 درجه، تغییر مقیاس تا 20 درصد، انتقال عرضی و ارتفاعی تا 10 درصد و برگردان افقی تصویر.

**2. پیش‌پردازش داده‌های آزمون:**

- نرمال‌سازی: تصاویر با مقیاس 255/1 نرمال‌سازی شدند تا پیکسل‌ها در محدوده [0, 1] قرار گیرند.

# مدل‌سازی با ResNet50V2

در این پروژه از شبکه‌ی عصبی ResNet50V2 استفاده شده است. ResNet50V2 یک شبکه‌ی عصبی عمیق و از پیش آموزش‌دیده شده با مجموعه داده‌های ImageNet است که شامل 50 لایه‌ی کانولوشنی می‌باشد. این مدل به دلیل عمق زیاد و ساختار خاص خود، قادر به استخراج ویژگی‌های بسیار پیچیده و مفید از تصاویر است.

ResNet مخفف "Residual Network" است که توسط محققان مایکروسافت برای حل مشکلات مربوط به آموزش شبکه‌های عمیق‌تر معرفی شده است.

ویژگی‌های کلیدی ResNet-50 v2:

**1. معماری Residual:**

- معماری ResNet از بلاک‌های "باقی‌مانده" (residual blocks) استفاده می‌کند. هر بلاک شامل یک اتصال کوتاه (shortcut connection) است که ورودی بلاک را مستقیماً به خروجی اضافه می‌کند. این ویژگی کمک می‌کند تا مشکلات مربوط به ناپدید شدن گرادیان‌ها (vanishing gradients) در شبکه‌های بسیار عمیق کاهش یابد.

**2. ساختار لایه‌ها:**

- ResNet-50 v2 شامل 50 لایه است که به صورت بلاک‌های باقی‌مانده سازمان‌دهی شده‌اند. این بلاک‌ها به صورت متوالی در عمق شبکه تکرار می‌شوند.

**3. بهبودهای نسخه v2:**

- در نسخه‌ی دوم ResNet (v2)، برخی بهبودهای معماری نسبت به نسخه‌ی اول (v1) اعمال شده است. یکی از این بهبودها، استفاده از "Batch Normalization" قبل از اعمال وزن‌ها در هر بلاک است که به تثبیت و سرعت بخشیدن به آموزش کمک می‌کند.

**4. عملکرد:**

- ResNet-50 v2، به دلیل عمق بالا و استفاده از بلاک‌های باقی‌مانده، دقت بسیار بالایی در تشخیص تصویر و دیگر وظایف دارد. همچنین، این شبکه توانسته است تا عمق زیادی بدون از دست دادن دقت به دلیل مشکلات گرادیان‌ها آموزش داده شود.

**مزایا:**

- دقت بالا: به دلیل استفاده از بلاک‌های باقی‌مانده و معماری بهینه، ResNet-50 v2 دقت بسیار بالایی در وظایف بینایی ماشین دارد.

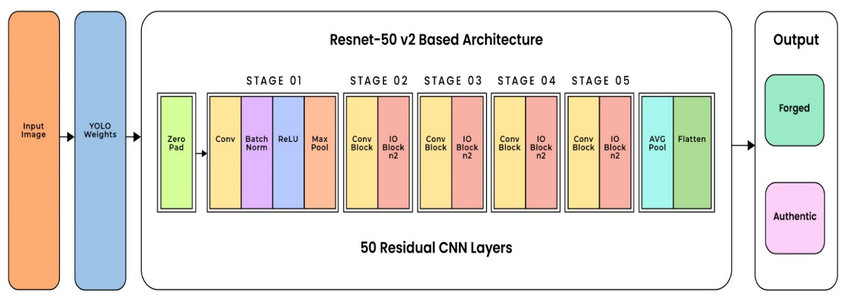
- پایداری در آموزش شبکه‌های عمیق: استفاده از اتصالات کوتاه به حل مشکلات مرتبط با گرادیان‌ها کمک می‌کند و امکان آموزش شبکه‌های بسیار عمیق‌تر را فراهم می‌کند.

**معایب:**

- پیچیدگی محاسباتی: به دلیل تعداد بالای لایه‌ها و عملیات‌های محاسباتی، ResNet-50 v2 نیاز به منابع محاسباتی زیادی دارد که ممکن است در برخی کاربردها محدودیت ایجاد کند.

ResNet-50 v2 یکی از معماری‌های محبوب و پرکاربرد در زمینه‌ی بینایی ماشین است که همچنان به عنوان یک استاندارد در بسیاری از پژوهش‌ها و کاربردهای صنعتی مورد استفاده قرار می‌گیرد.

در شکل ‏4‑1 معماری این شبکه نشان داده شده است.



شکل ‏4‑1 معماری شبکه‌ی ResNet-50 v2

مراحل پیاده‌سازی مدل به شرح زیر است:

**1. بارگذاری مدل ResNet50V2:**

- مدل ResNet50V2 با وزن‌های از پیش آموزش‌دیده شده بر روی مجموعه داده‌های ImageNet بارگذاری شد.

- لایه‌های بالایی مدل (به جز 5 لایه‌ی آخر) فریز شدند تا ویژگی‌های از پیش آموزش‌دیده شده حفظ شوند.

**2. ساخت مدل سفارشی:**

- افزودن لایه‌های Dropout، BatchNormalization، Flatten و Dense برای بهبود عملکرد مدل و جلوگیری از بیش‌برازش.

- خروجی نهایی با استفاده از یک لایه Dense با 7 نود (برای 7 کلاس احساسی) و فعال‌سازی softmax تولید شد.

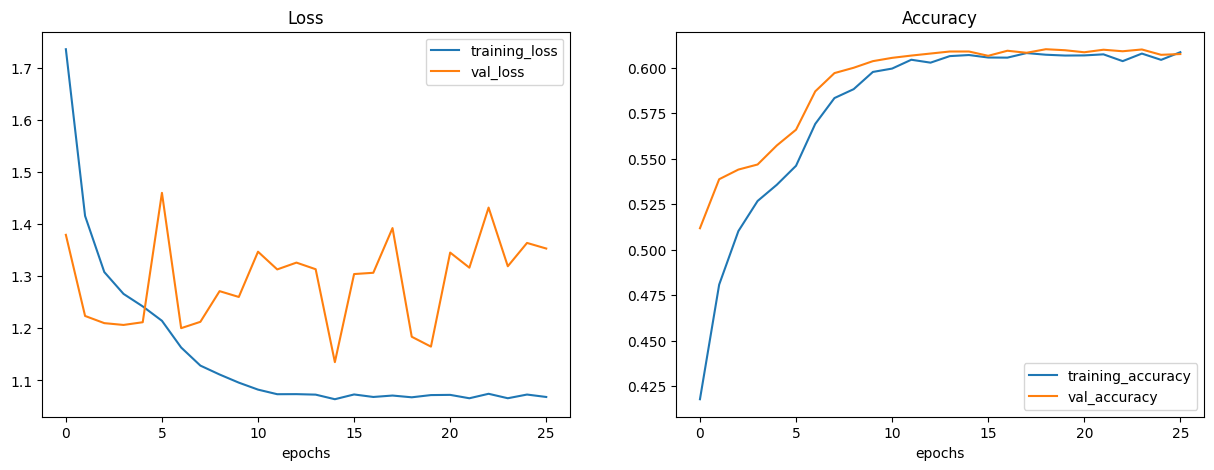
**3. آموزش مدل:**

- مدل با استفاده از optimizer آدم و تابع خطای categorical\_crossentropy کامپایل شد.

- از چندین callback برای بهبود فرآیند آموزش استفاده شد: Checkpoint برای ذخیره بهترین مدل، EarlyStopping برای توقف زودهنگام در صورت عدم بهبود دقت، و ReduceLROnPlateau برای کاهش نرخ یادگیری در صورت مشاهده بیش‌برازش.

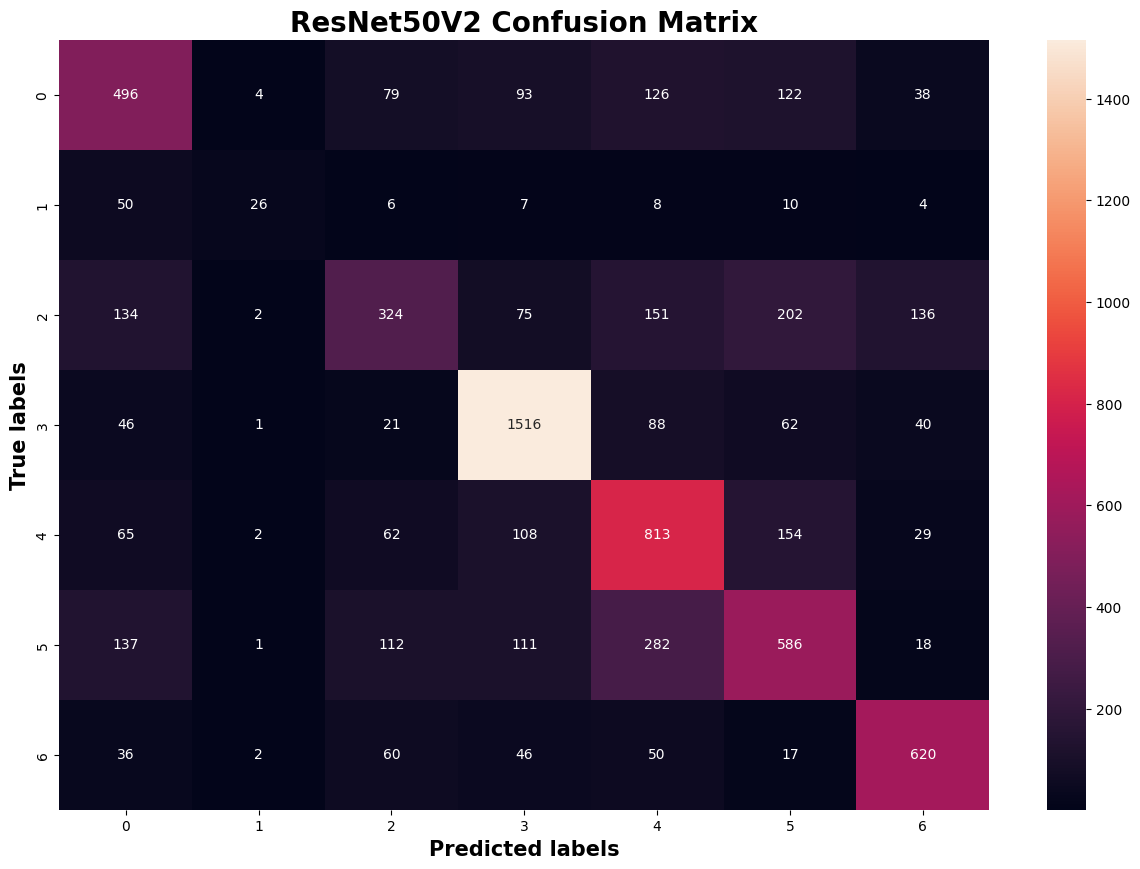
# نتایج

در شکل ‏5‑1 نمودار مربوط به مربوط به Loss و Accuracy نشان داده شده است.



شکل ‏5‑1 نمودار مربوط به loss و Accuracy

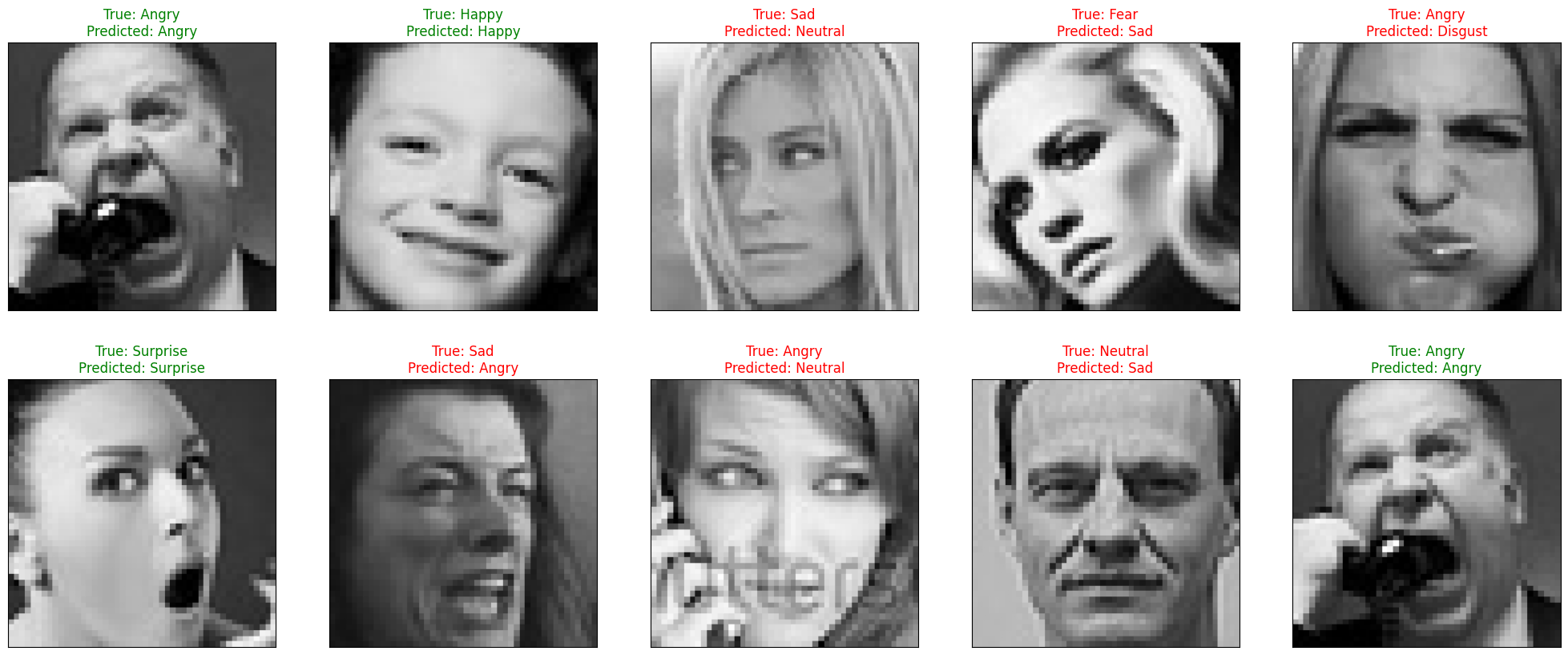
درشکل ‏5‑2 Confusion Matrix مربوط به خروجی نشان داده شده است.



شکل ‏5‑2 Confusion Matrix مربوط به خروجی

در نهایت صحت به دست آمده برابر با 61.03 درصد بود. علت پایین بودن صحت با وجود استفاده از شبکه‌ای قدرتمند، بالانس نبودن دادگان کلاس‌ها است به طوری که تعداد داده‌های مربوط به برخی کلاس‌ها تعداد کمی بودند و شبکه آن‌ها را به مقدار کافی یاد نگرفت. این مورد در Confusion Matrix نیز قابل مشاهده است. به طوری که برای مثال کلاس 1 به خوبی تشخیص داده نشده است.

در شکل ‏5‑3 نیز نمونه‌ای از تشخیص شبکه نشان داده است.



شکل ‏5‑3 لیبل تشخیص داده شده توسط شبکه برای تصاویر مختلف به همراه لیبل واقعی