

آماده‌سازی مجموعه داده‌ها

خواندن فایل‌های صوتی

مجموعه داده‌ها شامل فایل‌های صوتی با فرمت wav است. برای خواندن این فایل‌های صوتی و استخراج ویژگی‌های **Mel-Frequency Cepstral Coefficients (MFCCs)**، از کتابخانه **librosa** استفاده می‌کنیم که این ویژگی‌ها معمولاً برای پردازش صوتی به کار می‌روند.

<https://librosa.org/doc/latest/index.html>

استخراج ویژگی‌ها یا همان **Feature Extraction**

برای هر فایل صوتی، ویژگی‌های **MFCC** را استخراج می‌کنیم. این فرآیند شامل موارد زیر است:

- بارگذاری فایل صوتی
- محاسبه **MFCCs**، که نمایانگر طیف قدرت کوتاه‌مدت صدا هستند.
- میانگین‌گیری از **MFCCs** در طول زمان برای به دست آوردن یک بردار ویژگی با طول ثابت.

علاوه بر این، جنسیت گوینده را نیز به عنوان یک ویژگی با استخراج آن از نام فایل اضافه می‌کنیم.

تعریف و آموزش مدل

شبکه‌های **(Long Short-Term Memory (LSTM**

LSTM ها نوعی شبکه عصبی بازگشتی (**RNN**) هستند که برای مسائل پیش‌بینی دنباله مناسب‌اند. آن‌ها می‌توانند وابستگی‌های بلندمدت در داده‌های دنباله‌ای را ثبت کنند، که آن‌ها را برای وظایفی مانند تشخیص احساسات از صدا ایده‌آل می‌سازد.

جزئیات پیاده‌سازی:

معماری مدل:

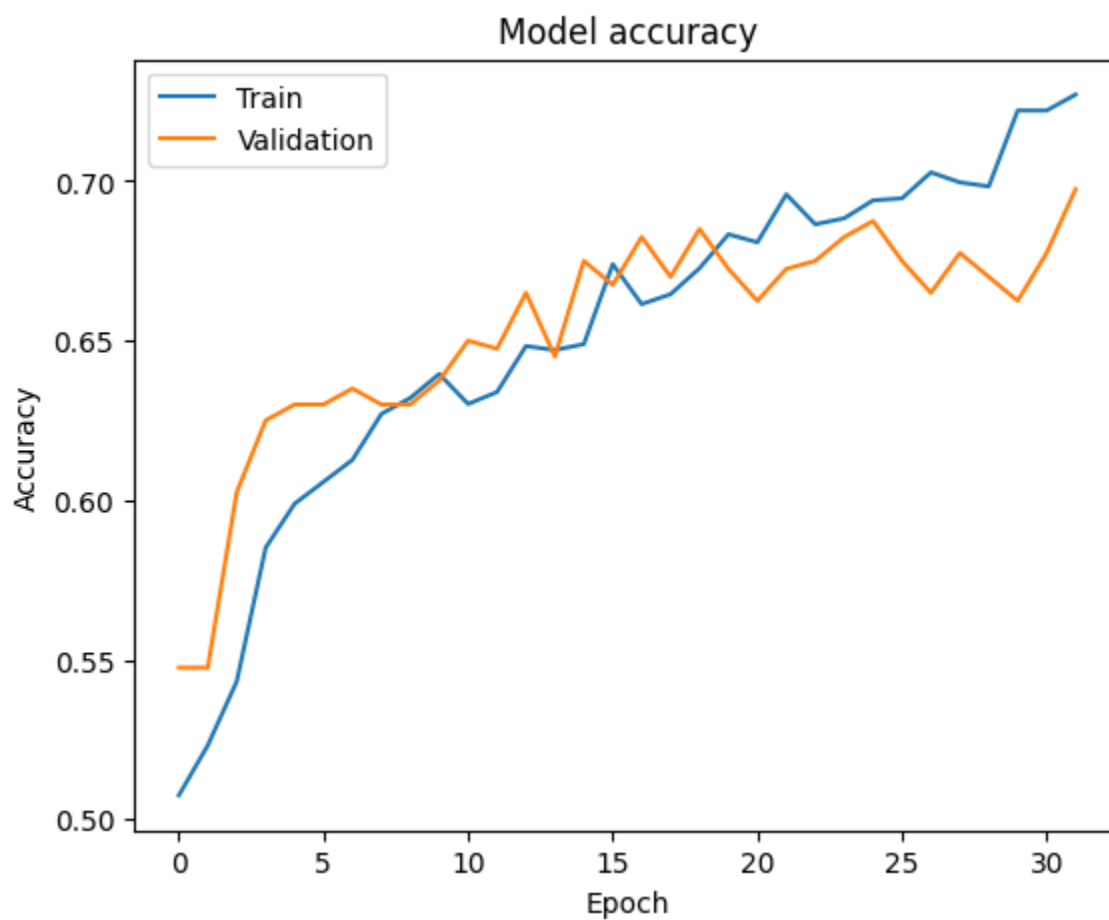
- مدل شامل دو لایه **LSTM** به دنبال لایه‌های **dropout** برای جلوگیری از overfitting است.
- لایه اول **LSTM** دنباله‌ها را بازمی‌گرداند، که به معنای خروجی گرفتن از کل دنباله پیش‌بینی‌ها برای هر گام زمانی است.
- لایه‌های **dropout** بعد از هر لایه **LSTM** اضافه می‌شوند تا بخشی از واحدهای ورودی به طور تصادفی به صفر تنظیم شوند، که به تنظیم مدل و کاهش overfitting کمک می‌کند.
- لایه نهایی یک لایه **dense** با تابع فعال‌سازی **softmax** است که احتمال‌ها را برای هر کلاس احساسات خروجی می‌دهد.

تنظیم مدل:

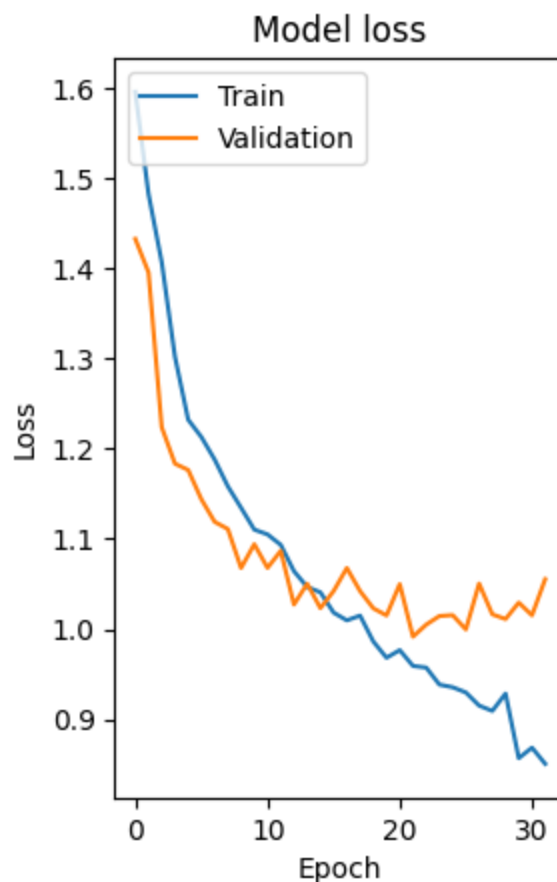
- **L2 Regularization**: به لایه‌های LSTM اعمال می‌شود تا وزن‌های بزرگ را جریمه کرده و مدل‌های ساده‌تری را تشویق کند.
- **Dropout**: نرخ dropout بر روی ۰/۶ تنظیم شده است تا سطح بالایی از تنظیم را تضمین کند.

Early Stopping:

- **early stopping** برای نظارت بر افت اعتبار استفاده می‌شود. آموزش متوقف می‌شود اگر افت اعتبار برای ۱۰ دوره متوالی بهبود نیابد. این از overfitting مدل به داده‌های آموزشی جلوگیری می‌کند.



در این پلات نسبت دقت مدل به تعداد **Epoch** ها نشان داده شده.
که بالا رفتن همزمان **Train** و **Validation** نشان دهنده درست **Fit** شدن مدل ما میباشد.
همچنین در نمودار زیر به بررسی **Loss** به تعداد **Epoch** ها پرداخته شده:



که نشان میدهد خطای مدل به میزان قابل توجهی پایین آمده.

نمودارهای تاریخچه آموزش

ما دقت و افت را برای هر دو مجموعه آموزشی و اعتبارسنجی رسم می‌کنیم تا عملکرد مدل را در طول دوره‌ها نظارت کنیم. این نمودارها به شناسایی **overfitting** کمک می‌کنند زمانی که عملکرد اعتبارسنجی شروع به کاهش می‌کند در حالی که عملکرد آموزشی همچنان بهبود می‌یابد.

آماده‌سازی داده‌های آزمایشی

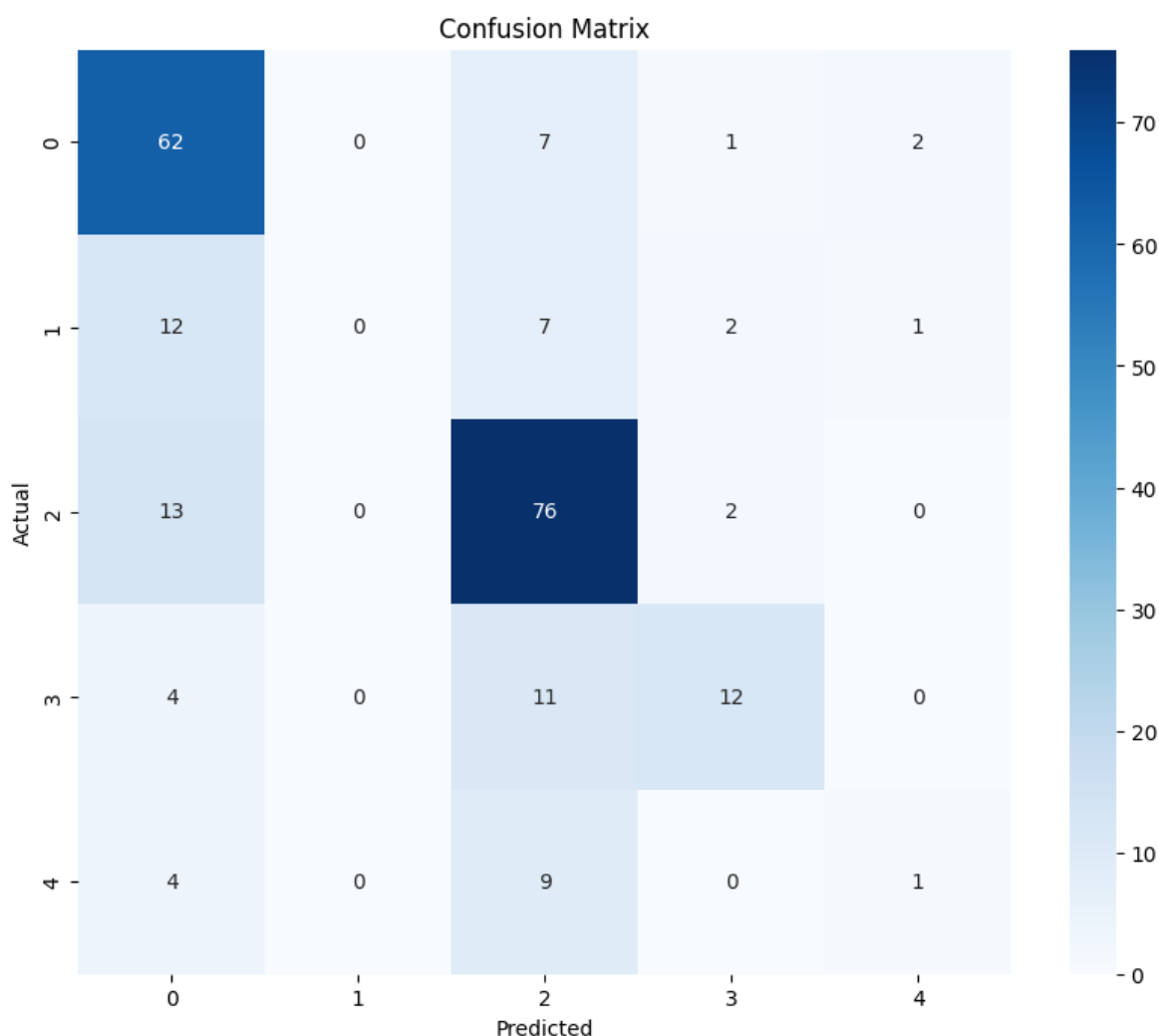
داده‌های آزمایشی به طور مشابه به داده‌های آموزشی آماده می‌شوند، با استخراج ویژگی‌های **MFCC** از هر فایل صوتی. برچسب‌های مجموعه آزمایشی در یک فایل **CSV** ارائه می‌شوند که ما آن را خوانده و برای ارزیابی مدل استفاده می‌کنیم.

معیارهای ارزیابی

- **Accuracy:** معیار اصلی استفاده شده برای ارزیابی مدل. این نسبت پیش‌بینی‌های درست به کل مشاهدات است.

```
8/8 [=====] - 0s 6ms/step - loss: 1.0381 - accuracy: 0.6726  
Test Accuracy: 0.6725663542747498
```

- **Confusion Matrix:** یک تحلیل دقیق از عملکرد مدل در کلاس‌های مختلف ارائه می‌دهد، و نشان می‌دهد تعداد true positive، false positive، true negative و false negative ها.

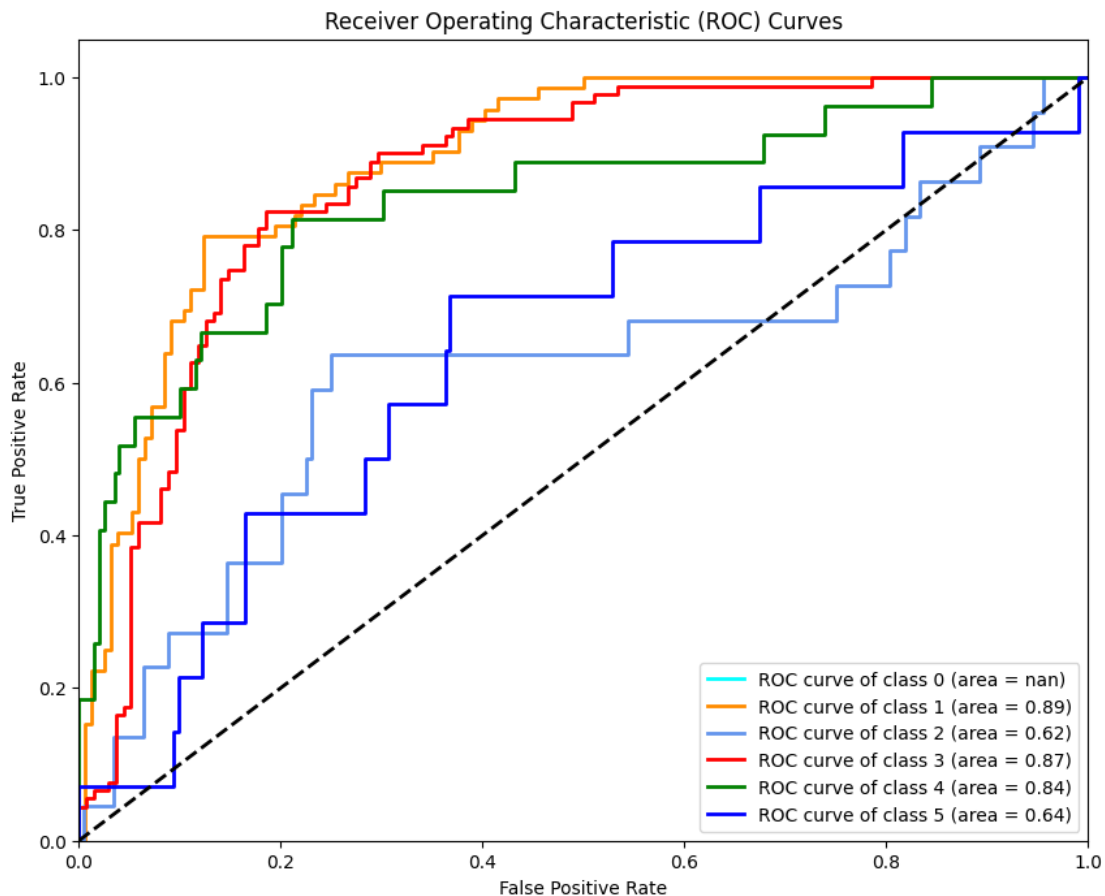


● **Classification Report**: شامل precision، recall و F1-score برای هر کلاس احساسی است.

- **Precision**: نسبت پیش‌بینی‌های مثبت درست به کل پیش‌بینی‌های مثبت.
- **Recall**: نسبت پیش‌بینی‌های مثبت درست به کل مشاهدات در کلاس واقعی.
- **F1-score**: میانگین وزنی precision و recall، که زمانی مفید است که نیاز به در نظر گرفتن هر دو معیار دارید.

ROC Curve and AUC:

● **ROC Curve**: نرخ true positive را در مقابل نرخ false positive در تنظیمات مختلف آستانه رسم می‌کند.



خلاصه مطلب:

- استخراج ویژگی: استفاده از **librosa** برای استخراج ویژگی‌های **MFCC** از فایل‌های صوتی.
- مدل **LSTM**: دو لایه **LSTM** با **dropout** و **L2 regularization** برای جلوگیری از **overfitting**.
- آموزش **early stopping**: بر اساس افت اعتبار برای توقف آموزش هنگامی که مدل شروع به **overfitting** می‌کند.
- ارزیابی: معیارهای جامع شامل دقت، ماتریس آشفتگی، گزارش طبقه‌بندی و نمودارهای ROC برای ارزیابی کامل عملکرد مدل.

نکات اضافی

- **LSTM** ها نسبت به **RNN** های ساده‌تر ترجیح داده می‌شوند زیرا مشکل **vanishing gradient** را کاهش می‌دهند و به آن‌ها اجازه می‌دهند وابستگی‌های بلندمدت در دنباله‌ها را ثبت کنند.
- تکنیک‌های تنظیم مانند **dropout** و **L2 regularization** در جلوگیری از **overfitting** بسیار مهم هستند، به خصوص با مدل‌های پیچیده‌ای مانند **LSTM** ها.
- **early stopping** به یافتن تعداد بهینه دوره‌های آموزشی کمک می‌کند، تعادل بین **overfitting** و **underfitting** را برقرار می‌کند.

این رویکرد ارزیابی و تفسیر قوی مدل تشخیص احساسات را تضمین می‌کند و بینش‌های ارزشمندی از عملکرد آن و زمینه‌های بهبود فراهم می‌کند