



دانشگاه صنعتی قوچان  
گروه مهندسی کامپیوتر

پروژه پایان ترم درس بینایی ماشین

عنوان پروژه

بازشناسی حالت چهره

Facial Expression Recognition

نام دانشجویان

امیرحسین رضائی پور ۹۷۱۳۳۱۰۵۷

ایمان خالقی ۹۷۱۳۳۱۰۴۳

استاد درس:

دکتر چم پور

زمستان ۱۴۰۰

## چکیده

حالت چهره یک نوع زبان بدن است که نقش مهمی در ارتباطات ایفا می‌کند. گاهی اوقات است که یک انسان سخن نمی‌گوید اما از حالت چهره او می‌توان اطلاعاتی درباره‌ی آن شخص دریافت کرد. این مسئله در خیلی جاها می‌تواند اهمیت داشته باشد به طور مثال در مشاوره‌ی یک روانشناس با یک بیمار و یا در ارتباط ربات با انسان (HRI). جملاتی که بیان شد اهمیت تشخیص حالت چهره را نشان می‌دهد.

در این گزارش سعی شده است تا با استفاده از تکنیک‌های یادگیری عمیق و شبکه‌های عصبی کانولوشنی به یک مدل برای تشخیص حالت چهره برسیم. برای پیدا کردن این مدل از دیتاست BU3DFE استفاده شده است و تلاش کرده‌ایم به ایجاد یک مدل که توانایی تشخیص ۷ کلاس مختلف از حالات چهره را داشته باشد که آن ۷ کلاس شامل خشم، غم، شادی، تعجب، عادی، نفرت و ترس می‌باشد.

فهرست مطالب

صفحه	عنوان
۱.....	فصل ۱- مقدمه.....
۱.....	۱-۱- پیشگفتار .....
۱.....	۱-۲- هدف از این الگو و دستورالعمل .....
۲.....	فصل ۲- پیش پردازش داده‌ها.....
Error! Bookmark not defined. ....	فصل ۳- شبکه‌ی کانولوشنی.....
۷.....	فصل ۴- نتایج ارزیابی .....
۱۰.....	فصل ۵- نتیجه گیری .....
۱۱.....	فصل ۶- منابع .....

## فصل ۱: مقدمه

### ۱-۱ پیشگفتار

در دهه‌های اخیر تکنولوژی باعث پیشرفت‌ها بسیاری شده است و توانسته مسائل بسیار زیادی را هم حل نماید. یکی از تکنولوژی‌هایی که نقش بسیار مهمی در پیشرفت و حل مسائل پیچیده و دشوار داشته است هوش مصنوعی می‌باشد. اما ما به طور مخصوص می‌خواهیم درباره‌ی یکی از زیرمجموعه‌های آن یعنی یادگیری عمیق صحبت کنیم. یادگیری عمیق که می‌توان گفت از سال ۲۰۱۲ به دستاوردهای بزرگی رسید، که این موفقیت‌ها را می‌توان در مسابقات *imagNet* و مقایسه نتایج قبل از ۲۰۱۲ و بعد از آن هم مشاهده کرد.

مسئله‌ای که ما سعی در حل کردن آن با یادگیری عمیق داریم، تشخیص حالت چهره انسان از روی تصویر در زوایای مختلف می‌باشد.

### ۱-۲ هدف از این مسئله

هدف از این مسئله ایجاد یک مدل است که توانایی این را داشته باشد که یک تصویر را بگیرد و تخمین بزند که این تصویر به کدام یک از ۷ کلاس خشم، غم، شادی، تعجب، عادی، نفرت یا ترس تعلق دارد.

برای این کار از شبکه‌های عصبی کانولوشنی استفاده کرده‌ایم. تصاویری که برای *train* کردن مدل استفاده کرده‌ایم مربوط به دیتاست *BU3DFE* می‌باشد. از آنجایی که داده‌های موجود در این دیتاست زیاد نیست، با استفاده از روش‌هایی اقدام به داده‌افزایی کرده‌ایم که در ادامه درباره آن‌ها توضیح داده خواهد شد.

## فصل ۲: پیش پردازش داده‌ها

همانطور که در قبل هم اشاره شد، تعداد داده‌های ما برای آموزش این شبکه کافی نیست از این رو با استفاده از روش‌هایی باید به داده افزایشی بپردازیم.

البته باید به این نکته هم توجه بشود که هر داده‌افزایی برای مسئله‌ی ما مفید نیست و باید سعی در پیدا کردن روش‌هایی کرد که در نهایت به بالاتر رفتن دقت مدل ما بینجامد.

یکی از روش‌های داده‌افزایی که برای این مسئله مفید است، تغییر در نورپردازی تصویر است. برای انجام این داده‌افزایی از تکنیک تصحیح گاما<sup>۱</sup> استفاده شده است.

با استفاده از فرمول زیر و تغییر پارامتر گاما می‌توانیم تغییر در نورپردازی تصویر را ایجاد کنیم.

$$y = \left(\frac{x}{255}\right)^{\frac{1}{\gamma}} \cdot 255$$

به نمونه‌های زیر دقت کنیم تا تاثیر پارامتر گاما بر روی تصویر را مشاهده نماییم.



ما با مقدار دهی اعداد ۰,۵ و ۲ به گاما توانستیم به ازای هر تصویر، ۲ تصویر را به دیتاست خود اضافه کنیم تا در روند یادگیری تاثیر مثبت داشته باشد.

یکی دیگر از داده‌افزایی‌هایی که استفاده کردیم، قرینه‌سازی گردن افقی و عمودی داده‌ها است که از داده‌افزایی‌های استاندارد است. نمونه‌ی این داده‌افزایی را در شکل زیر مشاهده می‌کنید.

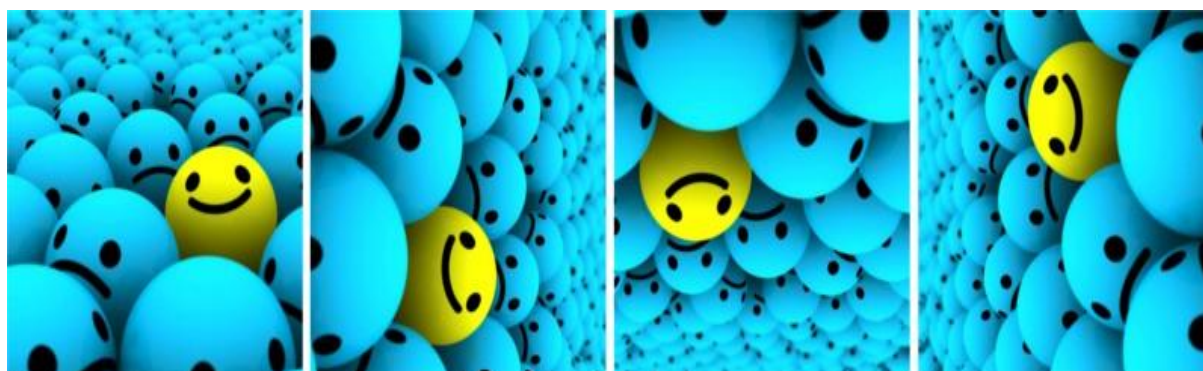
<sup>۱</sup>Data Augmentation

<sup>۲</sup>illuminations

<sup>۳</sup>gamma correction



و از دیگر روش‌های استفاده شده توسط ما برای این مسئله، چرخش در زوایای مختلف می‌باشد. در شکل زیر می‌توان به چند نمونه از چرخش را مشاهده کرد.



و بالاخره آخرین روش استفاده شده، برش تصویر می‌باشد. در این روش قسمتی از عکس را از تصویر جدا کرده و آن را به دیتاست اضافه خواهیم کرد. در شکل زیر نمونه‌ای از داده‌افزایی برش را مشاهده می‌کنید.



Flip

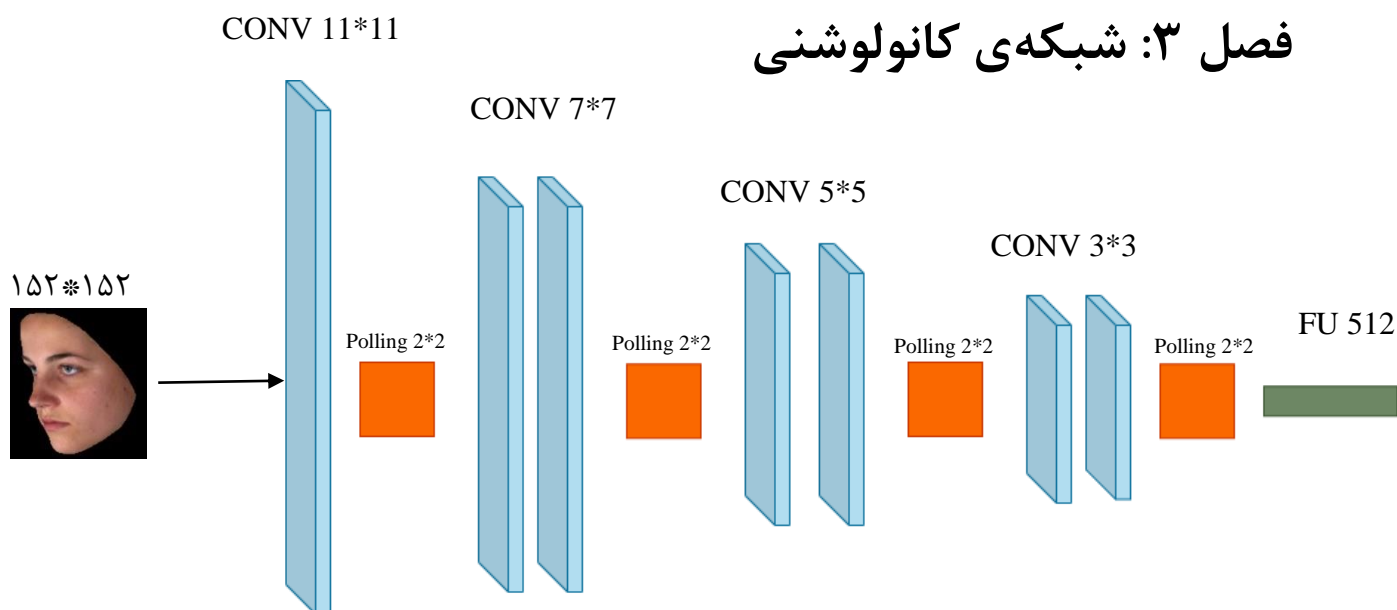
Rotate

Crop

هنگام داده افزایی باید بسیار مراقب بود که داده‌هایی که داریم به دیتاست اضافه می‌کنیم نباید موجب پدیده‌ی localization بشود زیرا در این صورت باعث overfit شدن مدل می‌شود و کارایی آن را پایین می‌آورد.

هرکدام از داده‌افزایی‌های انجام شده به نوعی به شبکه‌ی ما کمک می‌کند که هم از overfit جلوگیری بشود و هم با پیچیده تر شدن داده‌های train کمک می‌کند تا مدل در آینده و هنگام تست بتواند نتایج بهتری کسب کند، مثلاً تغییر در نورپردازی می‌تواند مدل ما را نسبت به نور مستقل کند.

### فصل ۳: شبکه‌ی کانولوشنی



همانطور که در شکل مشاهده می‌شود شبکه‌ی ما از ۴ لایه‌ی کانولوشن و یک لایه‌ی fully connected تشکیل شده است.

در اولین لایه تصویر ۱۵۹\*۱۵۹ پیکسلی ما که به صورت grayscale می‌باشد وارد شبکه می‌شود و با استفاده از ۳۲ فیلتر کانولوشن ۱۱\*۱۱ ویژگی‌هایی از آن استخراج می‌شود و سپس ابعاد خروجی با استفاده از max pooling به نصف کاهش پیدا کرده و سپس به لایه‌ی بعدی تحویل داده می‌شود. تصاویر ما دارای اطلاعات زیادی نیستند و حتی یک پس‌زمینه کاملاً سیاه در آن‌ها وجود دارد به همین علت از کانولوشن ۱۱\*۱۱ استفاده کرده‌ایم. در این لایه تعداد ویژگی‌های زیادی استخراج نمی‌شود و ویژگی‌های استخراج شده نیز سطح پایین می‌باشد اما مکان استخراج ویژگی‌ها بزرگ است زیرا داریم از کانولوشن ۱۱\*۱۱ استفاده می‌کنیم.

هم‌چنین در هر لایه، بعد از فیلترهای کانولوشن لایه‌ی batchNormalization آورده شده است که هدف از این کار نرمال سازی و جلوگیری از overfit شدن است.

به علاوه، بعد از لایه‌ی maxPooling یک لایه‌ی Dropout آورده شده است که این لایه هر بار یک تعداد از نورون‌ها به طور مثال ۲۰ درصد از نورون‌ها را خاموش می‌کند و با این کار نورون‌های باقی‌مانده باید جور نورون‌های از دست رفته را بکشند و سعی زیادتری در یادگیری یک مدل انجام بدهند و این کار به طور زیادی باعث جلوگیری از overfit شدن می‌شود.

ویژگی‌های استخراج شده در لایه‌ی اول بعد از گذشتن از لایه‌های batchNormalization, maxPooling و Dropout وارد لایه‌های کانولوشنی دوم می‌شوند که در آنجا دو لایه‌ی کانولوشنی ۷\*۷ وجود دارد. به نسبت لایه‌ی قبل تعداد فیلترها بالاتر رفته است زیرا ویژگی‌های سطح بالاتری را می‌توانیم استخراج



کنیم و ابعاد ماسک کانولوشن نیز کاهش پیدا کرده زیرا در این لایه ویژگی‌های پر اهمیت تری داریم و هم چنین ابعاد تصویر نیز کاهش پیدا کرده است.

خروجی این لایه نیز مانند لایه‌ی قبل وارد لایه‌های `batchNormaliztion`, `maxPooling` و `Dropout` می‌شود و سپس وارد لایه‌ی بعدی خواهد شد.

در قسمت بعدی مجدد سایز ماسک های کانولوشن از  $7 \times 7$  به  $5 \times 5$  کاهش پیدا میکند به دلایلی که در قسمت قبلی هم ذکر شد و خروجی ها پس از عبور از لایه‌های `batchNormaliztion`, `maxPooling` و `Dropout` تحویل قسمت بعد داده خواهند شد.

در قسمت بعدی سایز ماسک کانولوشن  $3 \times 3$  است زیرا دیگر عکس با عبور از چندین `maxPooling` کوچک شده است و ویژگی‌های قابل استخراج در این لایه بسیار بسیار با اهمیت هستند. پس از دو لایه‌ی کانولوشنی  $3 \times 3$  و عبور ویژگی‌های استخراج شده از `batchNormaliztion`, `maxPooling` و `Dropout` آنها را با استفاده از `Flatten` به صورت بردار در می‌آوریم و به دو لایه‌ی `Fully Conncted` وارد می‌کنیم و در انتها نیز با دستور :

```
model.add(Dense(num_classes, kernel_initializer='he_normal'))
```

آنها را به ۷ کلاس دسته بندی میکنیم، حالا وزن‌های شبکه‌ی ما توانایی تشخیص ۷ کلاس مختلف از حالت چهره را خواهند داشت.

## فصل ۴: نتایج ارزیابی

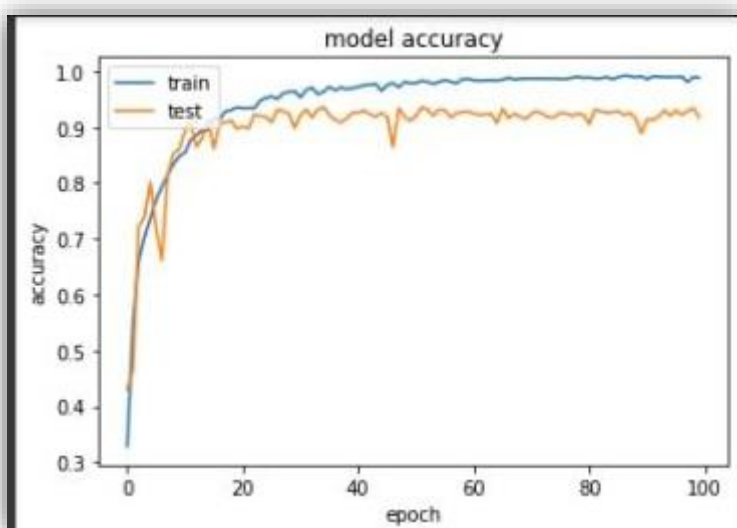
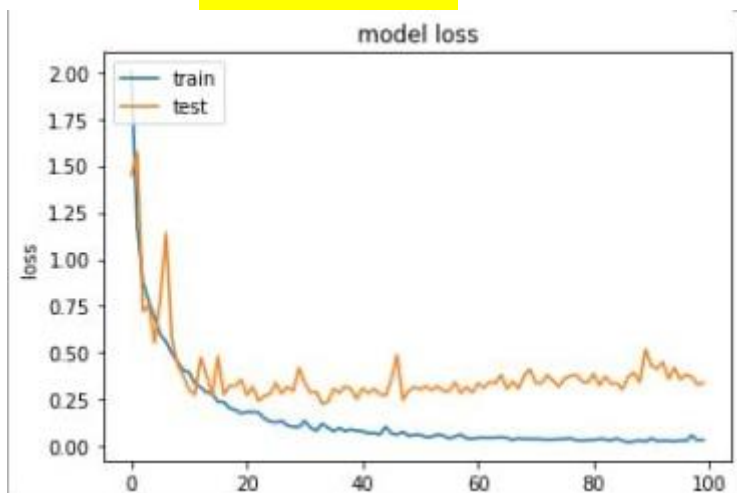
همانطور که گفته شد برای این مسئله ما از دیتاست BU3DFE استفاده کردیم و ۲۰٪ داده‌ها را برای ارزیابی و ۸۰٪ دیگر برای آموزش استفاده کردیم؛ و برای تقسیم داده‌ها از تابع زیر استفاده کردیم.

```
X_train,X_test,Y_train,Y_test = train_test_split(all_images,label, test_size=0.2 ,shuffle=True)
```

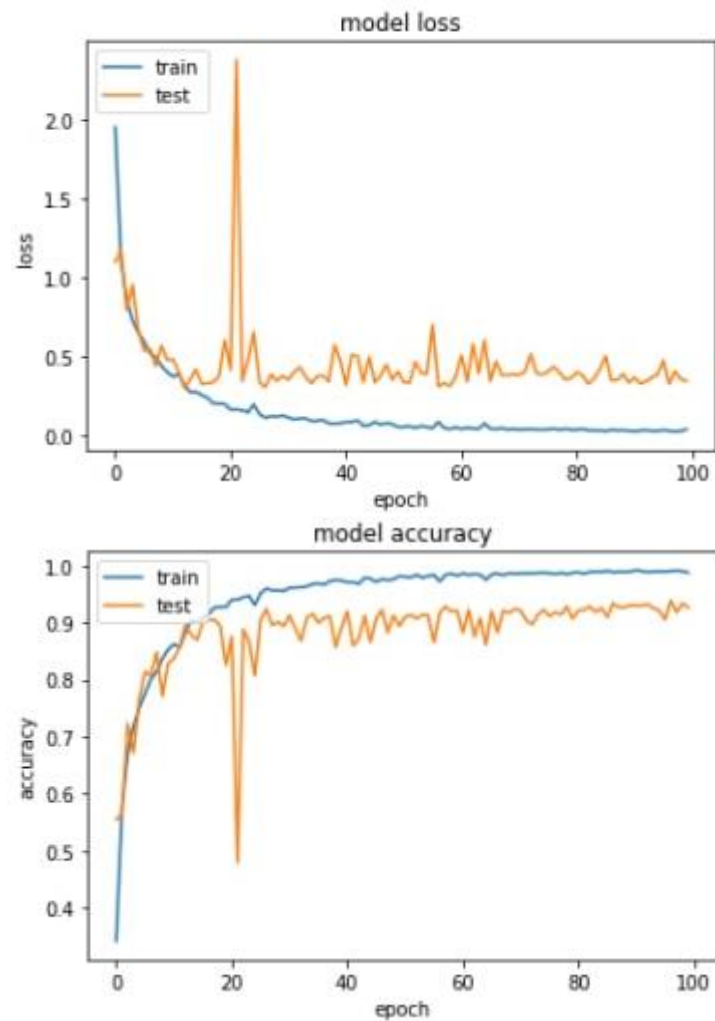
ویژگی این تابع این است قبل از جدا کردن داده‌ها shuffle می‌کند که باعث می‌شود هر دفعه داده‌های متفاوتی برای آموزش و آزمایش داشته باشیم.

ما ۳ مرتبه با این روش داده‌ها را در ۱۰۰ اپیاک آموزش و سپس ارزیابی کردیم و نتایج زیر بدست آمد:

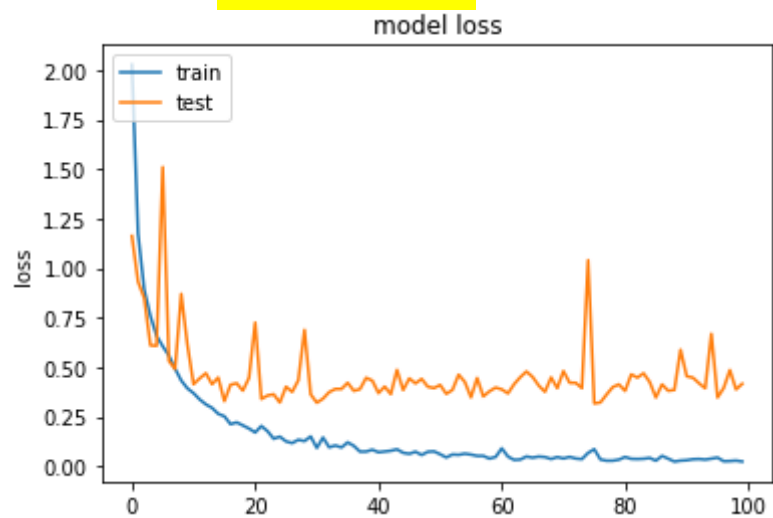
**Train ACC=98.80% , Test ACC=91.71%, Train Loss=0.035 , Test Loss=0.34(۱)**

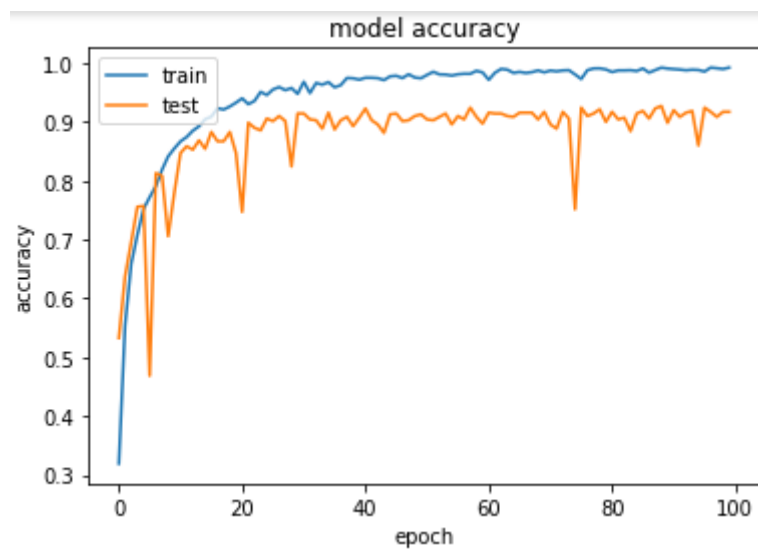


Train ACC=98.79% , Test ACC=92.71% , Train Loss=0.03 , Test Loss=0.34 (۲)

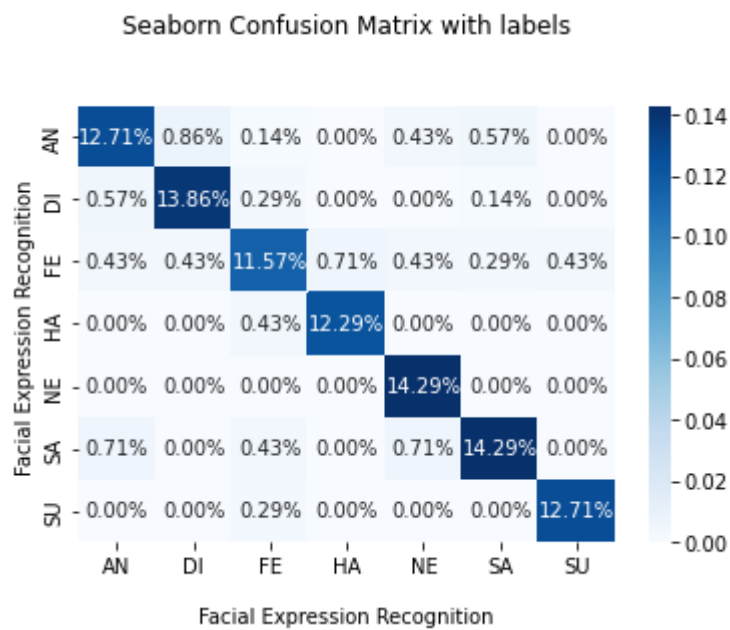


Train ACC=99.23% , Test ACC=91.71% , Train Loss=0.02 , Test Loss=0.41 (۳)





ماتریس اعوجاج برای ارزیابی مرتبه سوم



## فصل ۵: نتیجه گیری

در این مسئله سعی شد با استفاده یک شبکه عصبی عمیق که نسبت به شبکه های معروف دیگر پیچیدگی خاصی ندارد این مسئله حل شود؛ در واقع اگر ما از آن شبکه ها استفاده می کردیم به دلیل تعداد پارامترهای زیاد و نداشتن تعداد داده مناسب مسئله overfit می شد. و همچنین با داده افزایی مناسب توانستیم دقت خود را بالا ببریم و نتیجه ی نسبتا خوبی بگیریم.

## فصل ۶ : منابع

❖ **ilvia Ramis Guarinos,**

*Facial Detection and Expression Recognition applied to Social Robots,*  
July 2019,

Supervisors: Dr. Francisco J. Perales López, Dr. Jose Maria Buades Rubio  
and Dr. Jose Luis Lisani Roca,

**Universitat de les Illes Balears,**

Departament de Cie`ncies Matema`tiques i Informa`tica,  
UGIVIA Research group.

❖ <https://keras.io/>