



دانشگاه صنعتی اصفهان  
دانشکده مهندسی برق و کامپیوتر

# استفاده از رویکردهای ترکیبی یادگیری ماشین برای تشخیص اختلال اوتیسم با استفاده از داده های FMRI

نام و نام خانوادگی: سارا سلطانی گردفرامری  
اساتید راهنما: آقای دکتر حکیم داوودی و خانم دکتر مالکی

## ۱. مقدمه

### ○ تعریف مسئله

یک مسئله طبقه بندی افراد مبتلا به اوتیسم از افراد سالم با استفاده از داده های rs\_fmri از دیتاست ABIDE تعریف شده و در ادامه دو مدل مجزا با استراتژی های متفاوتی برای انتخاب ویژگی های مفید، آموزش مدل و در نهایت افزایش نمونه ها پیاده شده است.

با توجه به اهمیت اتصالات عملکردی در تحلیل داده های مغزی تعداد ویژگی ها بسیار زیاد است که نیاز به کاهش بعد می باشد. از طرف دیگر جمع آوری داده های تصویربرداری مغز هزینه بر می باشد پس در جهت بهبود عملکرد هر دو مدل استراتژی هایی در جهت افزایش داده به کار گرفته شد.

## ۲. پیاده سازی مدل اول

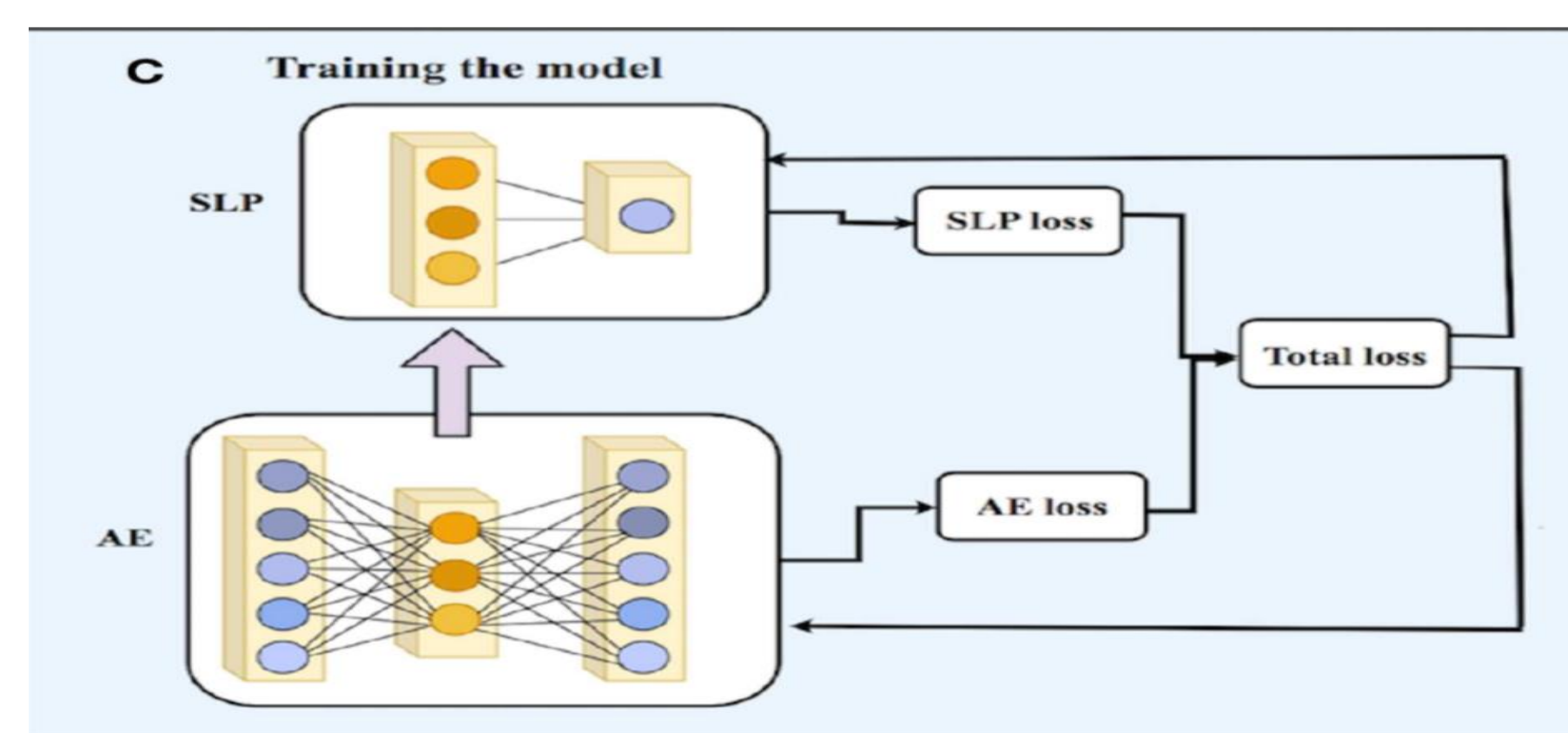
### ○ پیش پردازش اولیه داده ها

در این مرحله استراتژی های FILT\_GLOBAL و C\_PAC پیاده شده. سپس سه اطلس متفاوت روی نمونه ها اعمال می شود تا نتایج در نهایت با یکدیگر مقایسه شوند.

### ○ انتخاب ویژگی های مفید و آموزش مدل

- محاسبه ماتریس همبستگی برای تمامی نمونه ها و فشرده کردن قسمت مثلث بالایی
- محاسبه میانگین همبستگی ها برای هر ویژگی روی تمامی نمونه ها و انتخاب نهایی یک چهارم بزرگترین و یک چهارم کوچکترین ویژگی ها
- بکار گیری روش غیر خطی کاهش بعد به وسیله انکودر خودکار در کنار یک دیکودر با ساختار پیش رو

در جهت آموزش و طبقه بندی از پرسپترون تک لایه استفاده شده که انکودر خودکار و این پرسپترون به صورت همزمان آموزش داده می شوند.



### ○ افزایش تعداد داده با استفاده از درون یابی خطی

- انتخاب ۵ تا از نزدیکترین همسایگان هر نمونه با استفاده از معیار EROS

– این معیار شباهت بین دو سری زمانی چند متغیر (A, B) را براساس مقادیر ویژه و بردار های ویژه ماتریس های کوواریانس تمامی نمونه ها با استفاده از معادله زیر محاسبه می کند:

$$EROS(A, B, w) = \frac{\sum_{i=1}^n w_i |\langle a_i, b_i \rangle|}{\sum_{i=1}^n w_i |\cos \theta_i|}$$

- انتخاب تصادفی یک همسایه برای هر نمونه
- تولید داده جدید از داده اصلی و همسایه منتخب آن با استفاده از روش SMOTE

## ۳. پیاده سازی مدل دوم

### ○ انتخاب ویژگی های مفید و آموزش مدل

- استفاده از روش EXTRA\_TREE و شاخص جینی برای مشخص شدن با اهمیت ترین ویژگی ها
- انتخاب ۱۹۳۵ تا از بهترین ویژگی های به دست آمده از این روش
- آموزش مدل ماشین بردار پشتیبان

### ○ افزایش تعداد داده با استفاده از نسخه های متفاوت GAN

- استفاده از ساختار GAN
- تابع خطای مولد:

$$\mathcal{L}_G = -\mathbb{E}_{\hat{x} \sim p_{\text{fake}}} [\log D(\hat{x})]$$

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{x \sim p_{\text{data}}} [\|x - \hat{x}\|^2]$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_G + \mathcal{L}_{\text{rec}}$$

– تابع خطای متمایز کننده و متمایز کننده کد:

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_{\text{data}}} [\log D(x)] - \mathbb{E}_{\hat{x} \sim p_{\text{fake}}} [\log(1 - D(\hat{x}))]$$

$$\mathcal{L}_C = -\mathbb{E}_{z \sim p_{\text{latent}}} [\log C(z)] - \mathbb{E}_{z' \sim p_{\text{noise}}} [\log(1 - C(z'))]$$

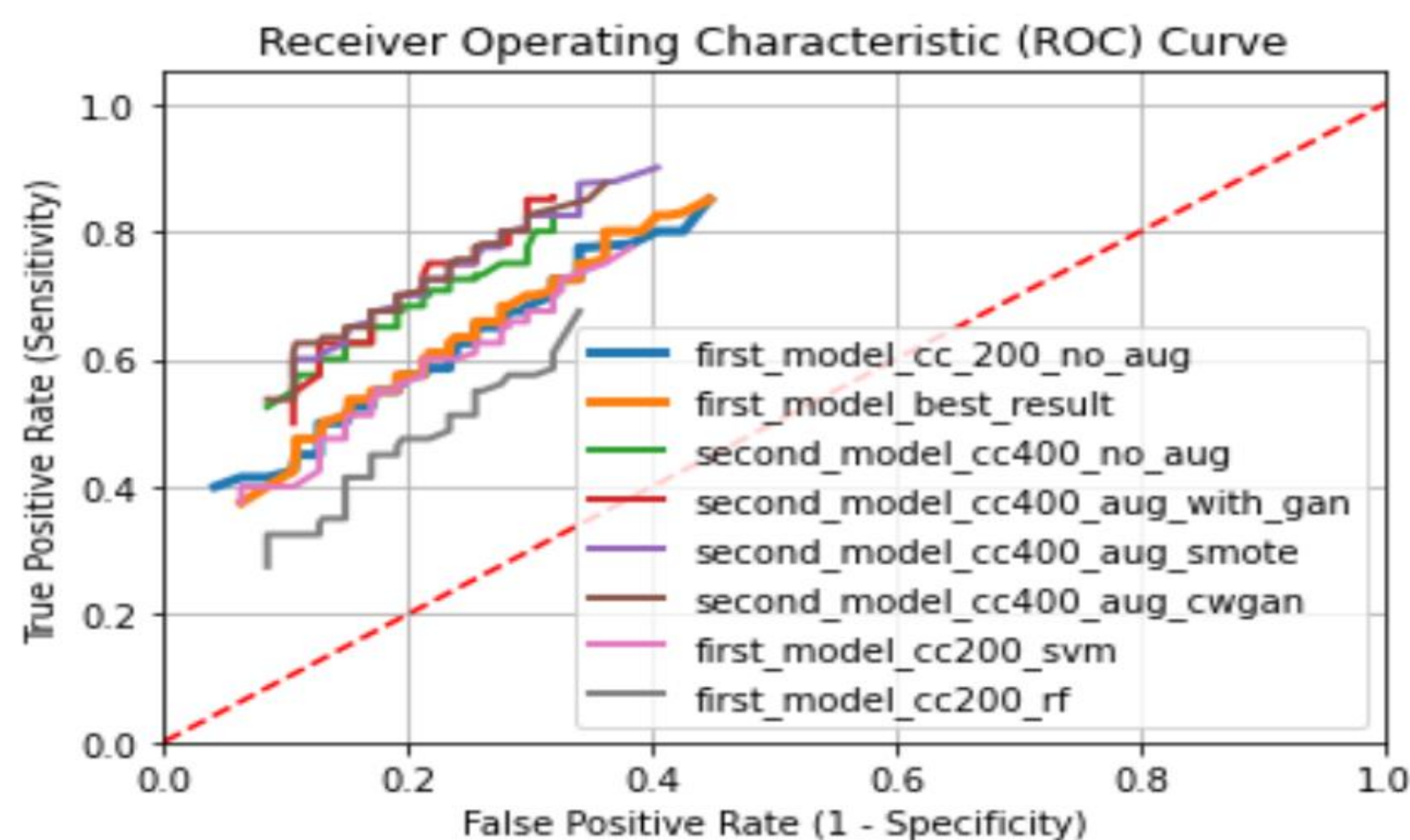
- استفاده از ساختار ترکیبی CONDITIONAL GAN, WGAN

$$\text{gradient penalty} = \mathbb{E}_{\hat{x} \sim p_{\hat{x}}} \left[ \left( \|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1 \right)^2 \right]$$

$$\mathcal{L}_D = -\mathbb{E}[D(x)] + \mathbb{E}[D(G(z))] + \lambda \cdot \text{Gradient Penalty}$$

$$\mathcal{L}_G = -\mathbb{E}[D(G(z))]$$

## ۴. نتایج نهایی و جمع بندی



models	Accuracy	Sensitivity	Specificity
Second model best result	75.4	70.9	79.4
first model best result	70.3	68	73