

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

موضوع

بررسی گراف های مغزی افراد به منظور یافتن ساختارهای مشترک

استاد راهنما

دکتر زینب مالکی

دکتر فرزانه شایق

دانشجو

ریحانه حلوائی ۹۶۲۶۷۹۳

ترم تابستان تحصیلی ۱۳۹۹-۴۰۰

فهرست

فصل اول مقدمه و تحقیقات مرتبط.....	۶
۱-۱- بیماری اوتیسم.....	۶
۲-۱- تحقیقات مرتبط.....	۶
فصل دوم مفاهیم و اطلاعات اولیه.....	۸
۱-۲- مغز.....	۸
۲-۲- بیماری اوتیسم.....	۸
۳-۲- تشخیص بیماری اوتیسم.....	۹
۴-۲- مدل کردن مغز به عنوان یک شبکه‌ی پیچیده.....	۹
۵-۲- ویژگی‌ها.....	۱۲
۶-۲- خوشه بندی با استفاده از مدل مخلوط گوسی.....	۱۵
۱-۶-۲- مدل مخلوط گوسی در مقایسه با <i>k-means</i>	۱۵
۲-۶-۲- عملکرد مدل مخلوط گوسی به اختصار.....	۱۵
۷-۲- تحلیل واریانس.....	۱۶
فصل سوم پیاده سازی و تحلیل.....	۱۸
۱-۳- معرفی مجموعه داده‌ی مورد استفاده.....	۱۸
۲-۳- روند کلی کار.....	۱۸
۳-۳- بررسی ویژگی‌ها با استفاده از تحلیل واریانس.....	۱۸
۴-۳- خوشه بندی با استفاده از ویژگی های خوب.....	۱۹
۵-۳- مقایسه ی نتایج حاصل از خوشه ها با برجسب های اصلی.....	۲۱
۱-۵-۳- مقایسه ی نتایج حاصل از خوشه ها با آستانه‌ای متفاوت.....	۲۸
۲-۵-۳- نتیجه گیری.....	۲۹
۶-۳- بررسی منطقه ای راس های منتخب.....	۳۰
فصل چهارم مقایسه و جمع بندی نهایی.....	۳۲
۱-۴- مقایسه.....	۳۲

۳۲ ۲-۴- جمع بندی نهایی

۳۴ فصل پنجم ضمیمه ها

۳۴ ۱-۵- جداول

۳۶ ۲-۵- تصاویر

۳۶ ۱-۲-۵- مقایسه خوشه ها (قبل و بعد از خوشه بندی با ویژگی های منتخب)

۴۰ فصل ششم مراجع

فصل اول

مقدمه و تحقیقات مرتبط

۱-۱- بیماری اوتیسم

اختلالات طیف اوتیسم^۱ به طور فزاینده ای شایع ترین اختلالات عصبی - رشدی هستند که با رفتارهای اجتماعی غیرمعمول مشخص می شوند، از جمله: نقص در بیانی رسا، نگرش و انعطاف پذیری ذهنی.

با این حال مشخص نیست که چگونه این تفاوت بین افراد سالم و دارای اختلال اوتیسم، در سطح شبکه منعکس می شود که در آن مغز به عنوان شبکه ای از صدها منطقه، که با یکدیگر ارتباط دارند، منعکس می شود.

پژوهش های قبلی صورت گرفته بر روی گراف های مغزی با استفاده از الگوریتم های کلاسیفیکیشن با هدف طبقه بندی آن ها بر روی افراد سالم و اوتیسم از دقت خوبی برخوردار نبوده؛ حال ما در این پژوهش به دنبال این قضیه هستیم که با انجام کلاسترینگ و بدون دانستن اطلاعاتی از قبیل سن و جنسیت افراد و صرفاً با بررسی بر روی خود مغز در نتیجه ی محدود کردن حیطه بررسی، ببینیم علت پایین بودن دقت چیست، آیا می توان با توجه به گسترده بودن طیف اختلالات این بیماری به دسته های متفاوتی از این بیماری دست یافت؟! ویژگی یا ویژگی های موثر در دسته بندی گراف های مغزی این افراد چیست؟

۱-۲- تحقیقات مرتبط

در مقالات زیادی با بررسی مغز، با استفاده از معیارهای شبکه های پیچیده، سعی به بهبود در تشخیص اختلالات اوتیسم داشته اند. در چند سال اخیر سعی به بررسی این بیماری با استفاده از روش های یادگیری عمیق نموده اند. در این روش ها ماتریس اتصال را به صورت یک تصویر در نظر می گیرند در حالی که ماهیت این ماتریس یک تصویر نیست بلکه گرافی است که دارای معیارهای شبکه های پیچیده است.

در پژوهش دیگری که رویکرد اصلی استفاده از گراف مغزی بوده است، با استخراج ویژگی های گرافی از این گراف و با استفاده از یادگیری بازنمایی سعی به بهینه سازی پردازش و کاهش پیچیدگی زمانی در شبکه یادگیری عمیق می نماید

¹ Autism Spectrum Disorders

تا با افزایش دقت بتواند به تشخیص افراد مبتلا به اختلالات اوتیسم کمک کند. روش یادگیری بازنمایی سعی می‌کند تا ویژگی‌های گراف و گره‌ها را حفظ کند.

همچنین برخی از پژوهش‌هایی که روی گراف‌های مغزی صورت گرفته، از مدل‌های یادگیری ماشین برای اندازه‌گیری دقت کلاسیفیکیشن استفاده شده است. در یکی از این پژوهش‌ها، باندهای فرکانسی مختلف اتصال‌های عملکردی مغز را به عنوان ویژگی استفاده کرده و با استفاده از SVM^2 کلاس بندی کرده است. و همچنین در سال ۲۰۱۵، از سه الگوریتم کلاس بندی SVM ، RF^3 و GBM^4 به این منظور استفاده کردند و در نهایت به دقت ۶۷ درصد رسیدند.

²Support vector machine

³ Random Forest

⁴ Gradient Boosting Machine

فصل دوم

مفاهیم و اطلاعات اولیه

۲-۱- مغز

درک سیستم مغزی به دلیل پیچیده بودن آن مشکل است و به همین دلیل تحقیقات پیرامون این رشته در دنیا نوپا بوده و به چند دهه اخیر برمی گردد. به جرات می توان مغز را پیچیده ترین شبکه‌ی جهان دانست که ساختاری در هم تنیده‌ای از عصب‌ها، آن را در احاطه‌ی خود درآورده‌اند و به لطفشان تک‌تک رفتارهای انسان و هر جاندار دیگری تحت کنترل آن در می آیند. چرا که شبکه‌ی مغز از میلیاردها نورون تشکیل شده که مطالعه‌ی ارتباطات میان آن‌ها و تشخیص عملکرد نورون‌ها در هر فعالیت برای شناخت مغز بسیار حائز اهمیت است. در واقع همین ارتباطات مختلف میان نورون‌ها است که می تواند منجر به بروز فعالیت خاصی در عملکرد انسان شود؛ این سیستم های پیچیده را می توان به عنوان مجموعه ای از "راس‌ها" (یعنی مناطق مغز) و "یال‌ها" (یعنی ارتباطات بین راس ها) توصیف و مدل کرد.

۲-۲- بیماری اوتیسم

به طور کلی اختلالات نافذ مربوط به رشد یا همان اختلالات فراگیر رشدی^۵ (PDD) به پنج گروه طبقه‌بندی می شوند: اوتیسم^۶، آسپرگر^۷، اختلالات از هم گسیختگی کودکی^۸، سندرم رت^۹ و اختلال فراگیر رشد که به گونه‌ای دیگر مشخص نشده^{۱۰} نام گذاری شده است در این بین سه گروه اوتیسم، آسپرگر و اختلال فراگیر رشد که به گونه‌ای دیگر مشخص نشده در حیطه اختلالات طیف اوتیسم^{۱۱} قرار می گیرند. شیوع این اختلال به شدت رو به افزایش است اما هنوز دلیل قطعی برای این اختلال بیان نشده است.

^۵ Pervasive Developmental Disorders

^۶ Autism

^۷ Asperger

^۸ Childhood disintegrative

^۹ Rett Syndrome

^{۱۰} Pervasive Developmental Disorders Not Otherwise Specified

^{۱۱} Autism Spectrum disorders

اوتیسم یا درخودماندگی یک اختلال رشد است و عموماً به اشتباه با کم‌توانی ذهنی یکسان پنداشته می‌شود که در سه سال اول زندگی ظاهر می‌گردد. این بیماری با تأثیر بر روی مغز کودک، رفتارهای اجتماعی و مهارت ارتباط برقرار کردن را مختل می‌کند. به کسانی که این اختلال را دارند "اوتیستیک" گفته می‌شود. کودک اوتیستیک حتی اگر کلام داشته باشد، نمی‌تواند درست ارتباط برقرار کند، آن‌ها ارتباط یک‌طرفه برقرار می‌کنند و به همین علت ارتباطشان موثر نیست؛ در این افراد حرکات تکراری دست‌زدن، پریدن پاسخ‌های غیرمعمول به افراد، دل‌بستگی به اشیاء و یا مقاومت در مقابل تغییر نیز دیده می‌شود. به نظر می‌رسد مساله ژنتیک مهم باشد اما عوامل محیطی و فاکتورهای ژنتیکی هر دو در بروز آن موثرند.

اختلال طیف اوتیسم بیان‌کننده یک طیف است که دارای سه ویژگی مشخص می‌باشد: نقص در ارتباط اجتماعی، انعطاف پذیر نبودن گفتار و رفتار، حرکات کلیشه‌ای؛ اگرچه اوتیسم به عنوان یک اختلال روانی مورد بحث قرار گرفته است اما همراه با دیگر اختلالاتی است که بیشتر فیزیکی یا سیستماتیک می‌باشد به واسطه گسترده بودن این طیف و همچنین تداخل آن با دیگر اختلالات بررسی و آنالیز آن بسیار مشکل خواهد بود.

۲-۳- تشخیص بیماری اوتیسم

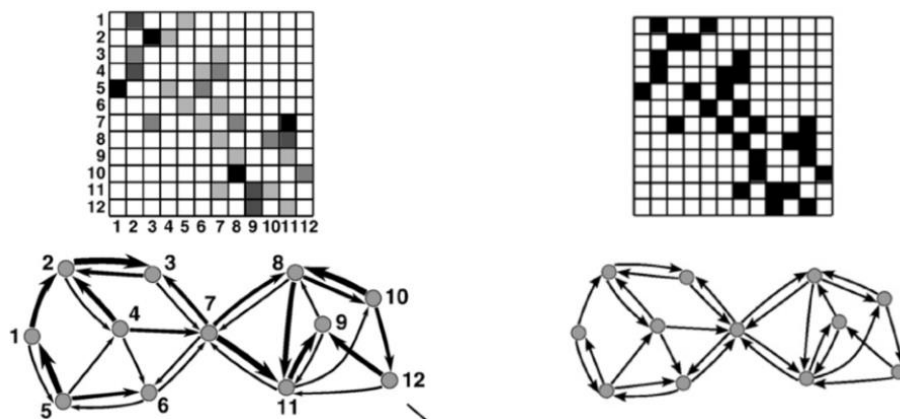
تشخیص افراد مبتلا به اختلالات اوتیسم یکی از مهم‌ترین اهداف در تحقیقات علوم‌شناختی است. موضوع مورد بحث در این زمینه، بررسی افراد مبتلا به اختلالات اوتیسم و نواحی مغزی‌ای که باعث این اختلالات می‌شوند است. یکی از روش‌هایی که اخیراً مورد توجه قرار گرفته ماتریس اتصالات مغزی است، این رویکرد باعث شد تا با استفاده از شبکه‌های پیچیده به تحلیل گراف‌های مغزی و مقایسه آن‌ها بپردازند. معمولاً ماتریس اتصال با استفاده از شبکه پیچیده تحلیل می‌شود. مقالات زیادی به بررسی مغز با استفاده از معیارهای شبکه‌های پیچیده به بهبود در تشخیص اختلال اوتیسم پرداخته‌اند. در این بخش به معرفی ماتریس اتصال مغزی یا کانکتوم می‌پردازیم که به ما کمک می‌کند درباره مغز به عنوان یک شبکه پیچیده، اطلاعات مورد نیاز را کسب کرده و عملیات مختلفی را بر روی این شبکه پیاده‌سازی کنیم. به یک نقشه‌ی جامع از اتصالات نورون‌ها در مغز **کانکتوم** می‌گویند که با نام ماتریس اتصال مغزی نیز شناخته می‌شود. به زبان جامع‌تر، یک کانکتوم تمام اتصالات نورونی درون مغز یک موجود زنده را شامل می‌شود. کانکتوم دارای دو بخش اصلی است: نورون‌ها و اتصالات نورونی. در این ماتریس $N \times N$ سطرها و ستون‌ها بیانگر N نورون و هر درایه‌ی a_{ij} نشان‌دهنده‌ی میزان ارتباط بین دو راس i و j به عنوان دو نورون در مغز است.

۲-۴- مدل کردن مغز به عنوان یک شبکه پیچیده

در ریاضیات یک شبکه یا گراف، مدل سازی ریاضی ای از مسئله های دنیای واقعی است که به صورت مجموعه هایی از رئوس و یال ها تعریف می شود؛ حال برای بررسی مغز به عنوان یک شبکه ی پیچیده و بررسی معیارهای مختلف بر روی آن، در اولین قدم نیاز است بدانیم رئوس و یال های گراف را به چه صورت تعریف می کنیم. رئوس معمولاً نشان دهنده ی نواحی مختلف مغز است که تعریف این نواحی بستگی به آن دارد که در چه سطحی در حال مطالعه ی مغز هستیم. در تعریف رئوس باید به این نکته توجه کرد که یک قسمت از مغز در دو ناحیه جای نگیرد؛ چرا که با در نظر گرفتن یک راس برای هر ناحیه، ممکن است هویت بخشی از مغز در دو راس تعریف شده باشد و این بررسی گراف را دچار اشکال خواهد کرد. از طرفی رئوس باید به گونه ای تعریف شوند که تمامی نواحی مغز را پوشش دهند تا عملکرد تمامی نواحی در مطالعه ی ما دخیل باشند. یال ها نشان دهنده ی اتصالات آناتومیک یا عملکردی بین رئوس بسته به نوع مجموعه داده ای است؛ اتصالات آناتومیک معمولاً مربوط به رشته های ماده ی سفید بین نواحی مختلف مغز است.

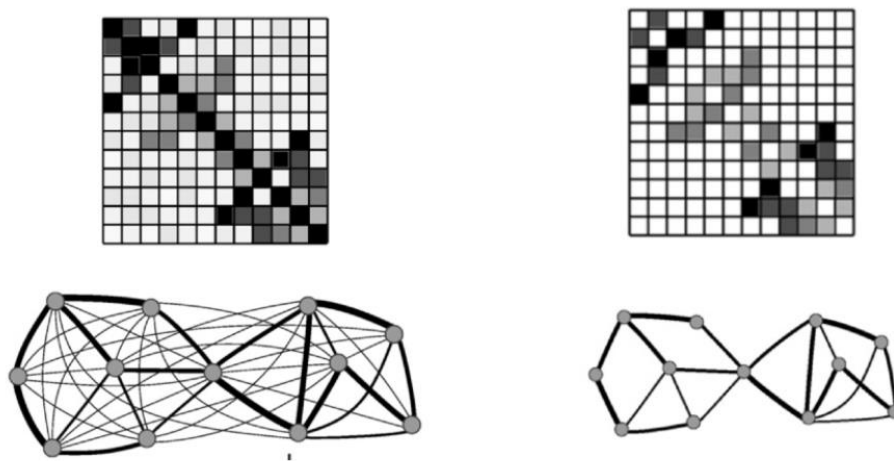
اتصالات عملکردی مغز، مربوط به میزان ارتباط قسمت های مختلف مغز در یک زمان حین انجام یک فعالیت است که ممکن است این قسمت ها با هم ارتباط آناتومیک نداشته باشند. نشان دادن ارتباطات بین نواحی مختلف مغز در گراف به صورت یال های وزن دار یا غیر وزن دار تعریف می شود. اگر گراف به صورت باینری (بدون یال های وزن دار) تعریف شود وجود یک یال نشان دهنده ی وجود ارتباط بین دو یا ناحیه ی مغز است. اگر گراف وزن دار باشد وزن هر یال نشان گر میزان قوی بودن ارتباط بین دو ناحیه یا چگالی سائز رشته های بین دو ناحیه در مغز است.

در تصویر ۱-۲ ماتریس های مختلف از مغز را مشاهده می کنیم. ماتریس اول که در عکس سمت چپ مشاهده می شود، مربوط به یک گراف وزن دار و جهت دار ساخته شده از ۱۲ ناحیه در مغز است. مثلاً در تصویر سمت راست، از گراف وزن دار موجود در تصویر اول وزن هایش را حذف کرده ایم و صرفاً در صورت وجود اتصال بین دو راس i و j ، درایه a_{ij} را یک قرار داده ایم. هم چنین در صورتی که بین دو راس i و j ، هیچ ارتباطی وجود نداشته باشد، درایه a_{ij} را صفر در نظر گرفته ایم.



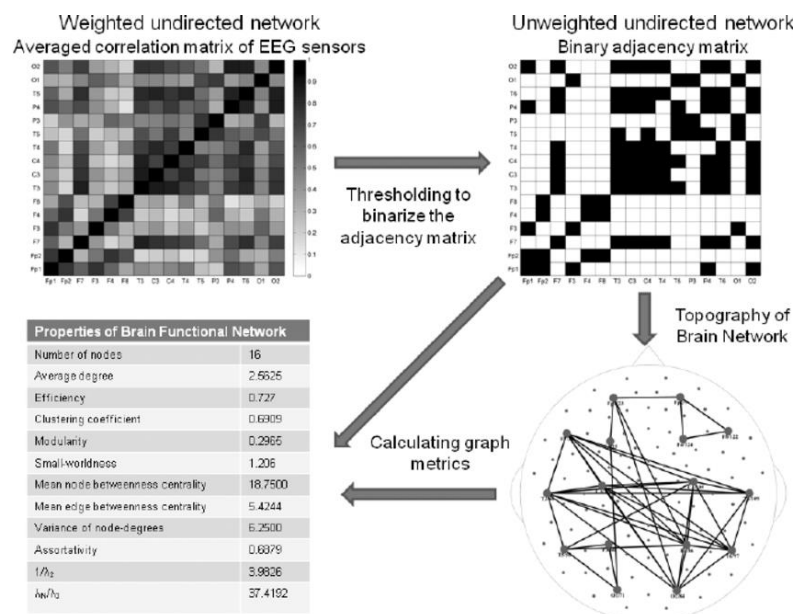
تصویر ۱-۲

در تصویر ۲-۲ ماتریس دیگری از همان نواحی در مغز را مشاهده می‌کنیم. این ماتریس یک ماتریس وزن‌دار و غیرجهت‌دار است. در اینجا هم سعی کرده‌ایم با در نظر گرفتن یک آستانه^{۱۲} برای وزن یال‌ها، یال‌های با وزن کم‌تر را حذف کرده و فقط یال‌هایی که در گراف حائز اهمیت هستند را نگه داریم.



تصویر ۲-۲

از روش بالا، در قسمت‌های بعدی این پژوهش برای حذف یال‌های با اهمیت کمتر در بررسی معیارها استفاده خواهیم کرد که به طور خلاصه می‌توانید در زیر روند کلی آن را مشاهده کنید:



تصویر ۳-۲

¹² threshold

۲-۵- ویژگی‌ها

به طور کلی برای هر گراف ویژگی‌های مختلفی تعریف می‌شود که هر کدام می‌تواند در تعریف معیارهایی برای بررسی عملکرد نواحی مغز کمک کند.

برخی از این ویژگی‌ها که در پروژه نیز به کار رفته است به شرح ذیل می‌باشد.

قدرت راس^{۱۳}: مجموع وزن یا های متصل به یک راس را قدرت آن راس می‌گوییم. در گراف‌های جهت‌دار قدرت ورودی راس مجموع وزن یال‌های ورودی آن است و قدرت خروجی راس مجموع وزن یال‌های خروجی است.

خروج از مرکز^{۱۴}: خروج از مرکز یک راس، بیشترین فاصله‌ی راس مذکور از دیگر راس‌های موجود در شبکه می‌باشد. به بزرگترین خروج از مرکز راس‌های گراف قطر و به کوچکترین خروج از مرکز راس‌های گراف شعاع گراف می‌گویند.

مثلث‌ها^{۱۵}: تعداد مثلث‌هایی است که راس مذکور در آن حضور دارد نسبت به کل مثلث‌هایی که می‌تواند در آن حضور داشته باشد.

ضریب خوشگی^{۱۶}: معیاری است که تمایل راس‌ها به ایجاد یک خوشه در یک گراف را اندازه می‌گیرد. برای هر راس در گراف، تعداد مثلث‌هایی که اطراف آن راس وجود دارد را ضریب خوشگی آن گفته می‌شود. به بیان بهتر همسایه‌هایی از گراف که خودشان با یکدیگر همسایه هستند را بررسی می‌کنیم. برای هر راس v در گراف G ضریب خوشگی به صورت زیر تعریف می‌شود:

$$c(v) = \frac{2T(v)}{\deg(v)(\deg(v) - 1)}$$

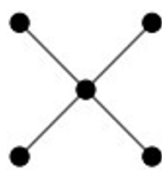
در رابطه‌ی فوق $T(v)$ تعداد یال‌هایی است که بین دو همسایه از راس v قرار دارد. اگر مقدار ضریب خوشگی برای یک راس برابر صفر شود به آن معناست که این گراف یک گراف ستاره ایست و اگر برابر یک شود به آن معناست که گراف یک خوشه است؛ یعنی هر راس از گراف به هر راس دیگر مسیری به طول یک دارد. در تصویر ۲-۴ یک گراف ستاره ای و یک خوشه را مشاهده می‌کنیم. با محاسبه ی ضریب خوشگی مطابق فرمول بالا، به ترتیب مقدار صفر و یک را خواهیم داشت:

¹³ Node strength

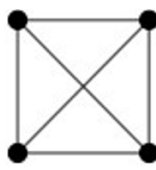
¹⁴ Eccentricity

¹⁵ triangles

¹⁶ clustering



$$\frac{2 * 0}{4 * (4 - 1)} = 0$$



$$\frac{2 * 3}{3 * (3 - 1)} = 1$$

تصویر ۲-۴

مرکزیت بر مبنای نزدیکی^{۱۷}: مقدار آن با معکوس میانگین طول کوتاه ترین مسیرهایی که از راس v به دیگر رئوس موجود در گراف برابر است. این معیار طبق فرمول زیر محاسبه می شود:

$$c_c(v) = \left[\frac{1}{n-1} \sum d(v, u) \right]^{-1} = \frac{n-1}{\sum d(v, u)}$$

که مقدار $d(v, u)$ کوتاه ترین مسیر بین u و v میباشد.

مرکزیت بینابینی^{۱۸}: مرکزیت بینابینی برای راس v ، نسبت تعداد کوتاه ترین مسیرهای بین هر دو راس s و t که از راس v میگذرد به تعداد کل کوتاه ترین مسیرهای بین دو راس s و t است. به بیان دیگر این معیار تعیین میکند که راس v در چه نسبتی از کوتاه ترین مسیرهای بین هر دو جفت راس در گراف ظاهر می شود. معمولاً راس هایی در گراف که به عنوان پلی بین بخشهای مجزای گراف هستند دارای مرکزیت بینابینی بالاتری نسبت به دیگر رئوس هستند.

برای محاسبه ی مرکزیت بینابینی هر راس v در گراف فرمول زیر را داریم:

$$c_B = \sum_{s, t \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)}$$

که $\sigma(s, t)$ تعداد کل مسیر های بین دو راس s و t میباشد و $\sigma(s, t|v)$ تعداد مسیر های بین دو راس s و t میباشد به شرط آنکه v در آن مسیر حضور داشته باشد.

مرکزیت بر مبنای درجه^{۱۹}:

مرکزیت بر مبنای درجه برای یک راس v تعداد راس هایی است که به آن متصل است.

$$c_D(v) = \frac{\deg(v)}{n-1}$$

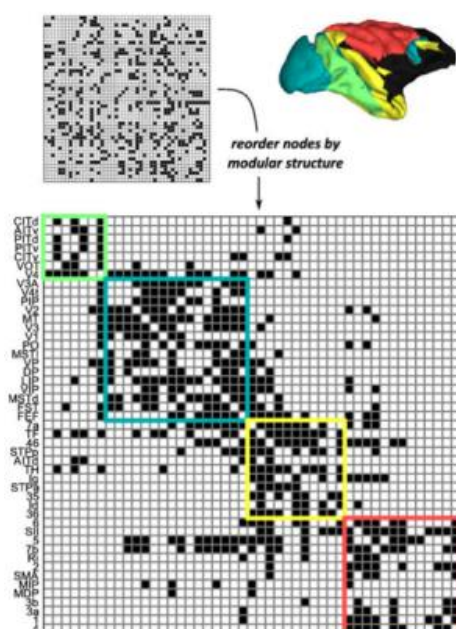
¹⁷ Closeness centrality

¹⁸ Betweenness centrality

¹⁹ Degree centrality

ماژولاریتی^{۲۰}: این معیار علاوه بر پیدا کردن ناحیه هایی که قابلیت تفکیک عملکرد دارند، سایز و ترکیب دقیق این نواحی را نیز تعیین می کند. با تقسیم رئوس به بخش های کوچک تر به گونه ای که داخل هر بخش تعداد ارتباطات بین آن ها ماکزیمم و بیرون از این بخش ها ارتباطات مینیمم باشد به این ترکیب ها دست پیدا می کنیم. به طور معمول معیارهای مورد بررسی ما به طور دقیق محاسبه شده و مقدار مشخصی را ارائه می دهند. اما برای پیدا کردن ماژولاریتی در گراف نیاز است از الگوریتم های بهینه سازی استفاده کنیم و جواب ما دقیق نخواهد بود. البته الگوریتم هایی که برای اینکار ارائه شده اند با دقت خوبی ماژول ها را تعیین می کنند.

در تصویر ۲-۵ نمونه ای از ماتریس مربوط به تصویر *fMRI* به دست آمده از قشر نوعی از میمون را مشاهده می کنیم. با به دست آوردن ماژول های قشر این میمون، ماتریس دوم را خواهیم داشت که داده ها برای هر ماژول از قشر قابل تفکیک است.



تصویر ۲-۵

بازده سراسری: میانگین معکوس طول کوتاه ترین مسیر بین هر جفت راس را بازده سراسری گوئیم. برعکس مشخصه ی طول مسیر، معیار بازده سراسری در گراف هایی که یال های کمتر بین رئوسشان وجود دارد محاسبه شده و معنا پیدا می کند. با توجه به این که در این گراف ها اگر بین دو راس یالی وجود نداشته باشد طول آن را بینهایت در نظر می گیریم، معکوس آن به صفر میل می کند و در نتیجه بازده برابر صفر را خواهد داشت با به طور کلی، مشخصه ی طول مسیر تحت اثر یال های با طول بیشتر در گراف و بازده سراسری تحت تاثیر یال هایی طول کمتر در گراف است.

۲-۶- خوشه بندی با استفاده از مدل مخلوط گاوسی^{۲۱}

مدل های مخلوط گاوسی یک مدل احتمالی برای نشان دادن زیرجمعیت هایی که به صورت نرمال توزیع شده اند، می باشد. مدل های ترکیبی به طور کلی نیازی ندارند که بدانند یک نقطه از داده به کدام زیرجمعیت تعلق دارد و به مدل اجازه می دهد تا زیرجمعیت ها را به طور خودکار بیاموزد. این مدل نوعی یادگیری بدون نظارت است.

۲-۶-۱- مدل مخلوط گاوسی در مقایسه با k -means

یکی از رایج ترین روش های خوشه بندی، الگوریتم خوشه بندی K -means است، که از یک روش تکرار شونده برای به روزرسانی پارامترهای هر یک از خوشه ها پیروی می کند. ما میانگین هر خوشه را محاسبه می کنیم که سپس با آن، میانگین هر خوشه را بدست آوریم و پس از آن، فاصله آن ها از هر نقطه داده را محاسبه کنیم. آنگاه الگوریتم، با شناسایی آن ها از طریق نزدیک ترین مرکز جرم، این نقاط داده را برچسب می زند. این روند تا رسیدن به یک معیار تغییر، تکرار می شود.

K -means یک الگوریتم خوشه بندی سخت است. با توجه به این، هر نقطه فقط به یک خوشه مرتبط می شود. به همین دلیل، یک عدم وجود احتمال داریم که می تواند به ما بگوید چند نقطه داده با یک خوشه خاص، مرتبط هستند. در نتیجه، از روش خوشه بندی نرم استفاده می کنیم. مدل های مخلوط گاوسی گزینه مناسبی برای این کار هستند.

۲-۶-۲- عملکرد مدل مخلوط گاوسی به اختصار

بسیاری از مجموعه داده ها می توانند به راحتی به کمک توزیع گاوسی مدل سازی شوند. بنابراین، می توان فرض کرد که خوشه هایی از توزیع های مختلف گاوسی هستند. ایده اصلی مدل این است که داده ها با مخلوطی از توزیع های گاوسی مدل سازی می شوند.

تابع چگالی احتمال یک بُعدی توزیع گاوسی به صورت زیر است:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

دو مقدار وجود دارد که مدل مخلوط گاوسی را پارامتری می کنند: وزن های مولفه، و واریانس ها / کواریانس ها. در یک مدل مخلوط گاوسی با K مولفه، μ_k میانگین مولفه k ام است. علاوه بر این، یک مورد تک متغیره، دارای متغیر واریانس σ_k می باشد در حالی که یک مورد چند متغیره، دارای یک ماتریس کوواریانس Σ_k خواهد بود. Φ_k تعریف وزن های مولفه مخلوط برای هر مولفه C_k است. یک قید $\sum_{i=1}^K \phi_i = 1$ وجود دارد، به گونه ای که احتمال کل باید به ۱ نرمال شود.

²¹ Gaussian mixture model

$$p(\vec{x}) = \sum_{i=1}^K \phi_i \mathcal{N}(\vec{x} | \vec{\mu}_i, \Sigma_i)$$

$$\mathcal{N}(\vec{x} | \vec{\mu}_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_i|}} \exp \left(-\frac{1}{2} (\vec{x} - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right)$$

$$\sum_{i=1}^K \phi_i = 1$$

۷-۲- تحلیل واریانس^{۲۲}

تحلیل واریانس و روش‌های تجزیه واریانس، یکی دسته از مدل‌های آماری هستند که قادرند اختلاف بین گروه‌ها یا دسته‌ها را بررسی کنند. در ساده‌ترین شکل، تحلیل واریانس می‌تواند به عنوان یک روش برای آزمون فرض مقایسه میانگین در بین چند جامعه مستقل به کار رود. این کار به عنوان یک جایگزین برای آزمون فرض با استفاده از آماره آزمون T است.

تحلیل واریانس در حالت کلاسیک راه حلی است که سه عمل زیر را همزمان انجام می‌دهد:

۱. تجزیه مجموع مربعات کل به مجموع مربعات اجزا حاصل از مدل خطی

۲. مقایسه میانگین مربعات، به کمک آماره و آزمون F

۳. آزمون پارامترهای مدل به منظور دستیابی به مدل آماری مناسب

فرض کنید قرار است در مورد یکسان بودن دو یا چند جامعه تحقیق کنید. یکی از شاخص‌های قابل استفاده برای بیان ویژگی‌های جامعه‌ها، می‌تواند میانگین باشد. با مقایسه میانگین و تشخیص برابری یا نابرابر بودن آن‌ها در بین جامعه‌ها، می‌توان رای به یکسان یا متفاوت بودن آن‌ها داد. بنابراین اگر یکی از میانگین‌ها با بقیه تفاوت داشته باشد، متوجه می‌شویم که جوامع مانند یکدیگر نیستند.

با توجه به این موضوع، می‌توان فرضیه‌های آزمون برای مقایسه میانگین k جامعه را به صورت زیر نوشت:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_1 : \text{there are some } \mu\text{'s not equal with others} \end{cases}$$

در اینجا فرض مقابل یا H1 بیان می‌کند که حداقل یکی از میانگین‌ها با بقیه تفاوت دارد. می‌دانیم که احتمال خطای نوع اول برای مسئله اصلی آزمون به صورت زیر نوشته می‌شود:

$$\alpha_t = P(\text{Error Type I}) = P(\text{Reject } H_0 | H_0 \text{ is true})$$

این عبارت به معنی احتمال رد فرض صفر به شرط صحیح بودن آن است. صحیح بودن فرض صفر بیانگر برابر بودن میانگین‌ها با یکدیگر خواهد بود. برای انجام این بررسی ممکن است از آزمون مقایسه میانگین در بین دو جامعه استفاده کنیم و به صورت ترکیب‌های دوتایی مسئله را تحلیل کنیم.

تحلیل واریانس در مسائل مربوط به آزمون میانگین چند جامعه، بر اساس تجزیه «پراکندگی کل»^{۲۳} به «پراکندگی بین گروه‌ها»^{۲۴} و «پراکندگی درون گروه‌ها»^{۲۵} صورت می‌پذیرد. بنابراین اگر پراکندگی کل را با «مجموع مربعات کل»^{۲۶}، پراکندگی بین گروهی را با «مجموع مربعات بین گروه‌ها»^{۲۷} و پراکندگی درون گروهی را با «مجموع مربعات درون گروهی»^{۲۸} نشان دهیم، خواهیم داشت:

$$SST = SSB + SSW$$

بر این اساس، با توجه به حضور یک متغیر تاثیرگذار در جامعه‌ها، محاسبات مربوط به جدول تحلیل واریانس برای آزمون مقایسه میانگین چند جامعه طبق جدول «تحلیل واریانس یک طرفه»^{۲۹} انجام می‌شود. در زیر یک نمونه از چنین جدولی قابل رویت است:

منشاء تغییرات	درجه آزادی	مجموع مربعات	میانگین مربعات	آماره F
بین گروه‌ها	$k - 1$	SSB	$MSB = \frac{SSB}{k - 1}$	$F = \frac{MSB}{MSW}$
درون گروه‌ها	$n - k$	SSW	$MSW = \frac{SSW}{n - k}$	
کل	$n - 1$	SST		

جدول ۱-۲

در این جدول k تعداد گروه‌ها یا جامعه‌ها و n نیز تعداد مشاهدات است.

متغیری که باعث اختلاف در جامعه‌ها می‌شود را گاهی «عامل»^{۳۰} نیز می‌نامند. در نتیجه اگر فرض صفر در آزمون فرض میانگین چند جامعه رد شود، می‌توان گفت که متغیر عامل در تغییر میانگین جامعه موثر است.

در صورتی که مقدار p -value کمتر از ۰.۰۵ باشد بدین معناست که فرض صفر رد شده و پراکندگی درون گروه‌ها کمتر از پراکندگی بین گروه‌ها خواهد بود.

²³ Total Variation

²⁴ Variation between groups

²⁵ Variation within groups

²⁶ Total Sum of Squares- SST

²⁷ Between Sum of Squares- SSB

²⁸ Within sum of squares- SSW

²⁹ One Way Anova

³⁰ Factor

فصل سوم

پیاده سازی و تحلیل

۳-۱- معرفی مجموعه داده‌ی مورد استفاده

مجموعه داده‌ی مورد استفاده *UCLA Autism* می‌باشد که شامل ۷۹ گراف مغزی با اندازه ماتریس 264×264 می‌باشد. تعداد بیماران مبتلا به اوتیسم در این مجموعه داده ۴۲ و تعداد افراد سالم ۳۷ نفر می‌باشد. محدوده‌ی سنی افراد در این مجموعه داده از ۷ تا ۱۸ سال می‌باشد.

۳-۲- روند کلی کار

در ابتدای کار، ماتریس مجاورت گراف را با آستانه‌ی ۰.۴ به یک ماتریس باینری تبدیل کرده و سپس از گراف مغزی هر فرد، ویژگی‌های ذکر شده در فصل قبل استخراج می‌شوند. در ویژگی‌هایی که برای هر راس تعریف شده‌اند، مقادیر ماکسیمم، مینیمم، میانه، مقدار مربوط به چارک اول و سوم و میانگین انتخاب می‌شوند.

در نهایت با استفاده از مقادیر و ویژگی‌های ذکر شده، گراف‌های مغزی را به کمک مدل ترکیبی گوسی در دو، سه و چهار دسته، خوشه‌بندی می‌شوند.

در ادامه با استفاده از تحلیل واریانس، ویژگی‌های خوب^{۳۱} انتخاب شده و بار دیگر خوشه‌بندی انجام می‌شود و در نهایت گراف‌های مغزی هر خوشه با استفاده از ویژگی‌های جدید تحلیل و ارزیابی می‌شوند.

۳-۳- بررسی ویژگی‌ها با استفاده از تحلیل واریانس

^{۳۱} ویژگی‌هایی که مقدار p-value آن‌ها در تحلیل واریانس کمتر از ۰.۰۵ باشد.

همان‌طور که در قسمت قبل ذکر شد، ابتدا ویژگی‌ها با استفاده از تحلیل واریانس بررسی می‌شوند. نتایج حاصل از این تحلیل و راهنمای آن در جدول ۵-۱ قابل مشاهده است.

در اینجا قسمتی از این جدول برای توضیحات بیشتر آورده شده است:

Features			2 clusters	3 clusters	4 clusters	Real labels
Clustering	18	Max	0.00020321	2.4254e-05	6.3371e-09	0.03906564
	19	Chart3	1.7359e-10	2.7442e-14	1.1882e-24	0.03664204
	20	Med	4.5937e-11	2.5221e-15	3.6337e-25	0.01836195
	21	Mean	4.3374e-12	2.9899e-16	2.8585e-25	0.01678299
	22	Chart1	6.3767e-12	1.5440e-15	5.1937e-19	0.00853054
	23	Min	0.31523237	0.05324934	0.03038160	0.16504208
Betweenness centrality	24	Max	0.00077406	4.7975e-05	0.00024072	0.11308404
	25	Chart3	9.8348e-13	3.9600e-13	3.6934e-13	0.03121171
	26	Med	4.4255e-14	9.4242e-12	5.9235e-13	0.03081895
	27	Mean	2.2377e-15	8.1552e-17	2.2771e-19	0.01545228
	28	Chart1	6.4447e-06	9.5050e-07	7.1972e-11	0.04962542
	29	Min	0.35683181	0.06787247	0.08268949	0.09328671

جدول ۳-۱

همان‌طور که در راهنمای جدول نیز گفته شده است، رنگ سبز (در ستون ویژگی‌ها) نشان دهنده‌ی این است که این ویژگی، هم برای خوشه‌های حاصل از خوشه‌بندی و هم برای برجسب‌های اصلی، ویژگی خوبی می‌باشد؛ یعنی برای مقادیر این ویژگی واریانس بین گروه‌ها نسبت به واریانس درون گروه‌ها به طور معناداری زیاد می‌باشد.

و همین‌طور رنگ نارنجی نیز به این معنا می‌باشد ولی برای برجسب‌های اصلی اینگونه نیست.

در اینجا مقدار ماکسیمم مرکزیت بینایی، از مقادیری می‌باشد که در خوشه‌بندی‌ها خوب عمل می‌کند ولی درواقع ویژگی مناسبی برای مقایسه بین گروه‌های سالم و ناسالم نمی‌باشد. به بیانی دیگر راسی که بیشترین مقدار مرکزیت بینایی را دارد، یعنی راسی که بیشترین تاثیر را در جابه‌جا کردن اطلاعات بین منطقه‌های مختلف مغز را دارد، معیار مناسبی برای جدا کردن دو گروه سالم و ناسالم نمی‌باشد.

در ادامه، ویژگی‌های سفید (که در خوشه‌بندی‌ها و برجسب‌های اصلی خوب عمل نکرده است) و ویژگی‌های نارنجی (که در خوشه‌بندی‌ها خوب عمل کرده است ولی درواقع ویژگی خوبی برای جدا کردن دو گروه سالم و ناسالم نمی‌باشد) از ویژگی‌ها حذف می‌شوند و بار دیگر خوشه‌بندی انجام می‌شود.

۳-۴- خوشه‌بندی با استفاده از ویژگی‌های خوب

در این قسمت نحوه ی تغییر خوشه‌بندی‌ها، با کل ویژگی‌ها و ویژگی‌های خوب نشان داده می‌شوند.

ویژگی‌های خوب، ویژگی‌هایی هستند که با آستانه ی ۰.۰۵ روی p -value های حاصل از تحلیل واریانس انتخاب شده‌اند.

نتایج با جزئیات بیشتر از چگونگی تغییر کلاسترها در فصل ضمیمه ها آورده شده‌است.

■ خوشه بندی بر اساس دو دسته:

در تصویر ۱-۳ نشان داده شده‌است که در خوشه ی اول، بعد از خوشه‌بندی با ویژگی‌های خوب، درصد افراد سالم افزایش میابد. و در خوشه ی دوم درصد افراد ناسالم کمی افزایش میابد.

■ خوشه بندی بر اساس سه دسته:

در تصویر ۱-۳ نشان داده شده‌است که در خوشه ی اول، بعد از خوشه‌بندی با ویژگی‌های خوب، درصد افراد سالم افزایش میابد. و در خوشه ی سوم نیز درصد افراد ناسالم کمی افزایش میابد.

■ خوشه بندی بر اساس چهار دسته:

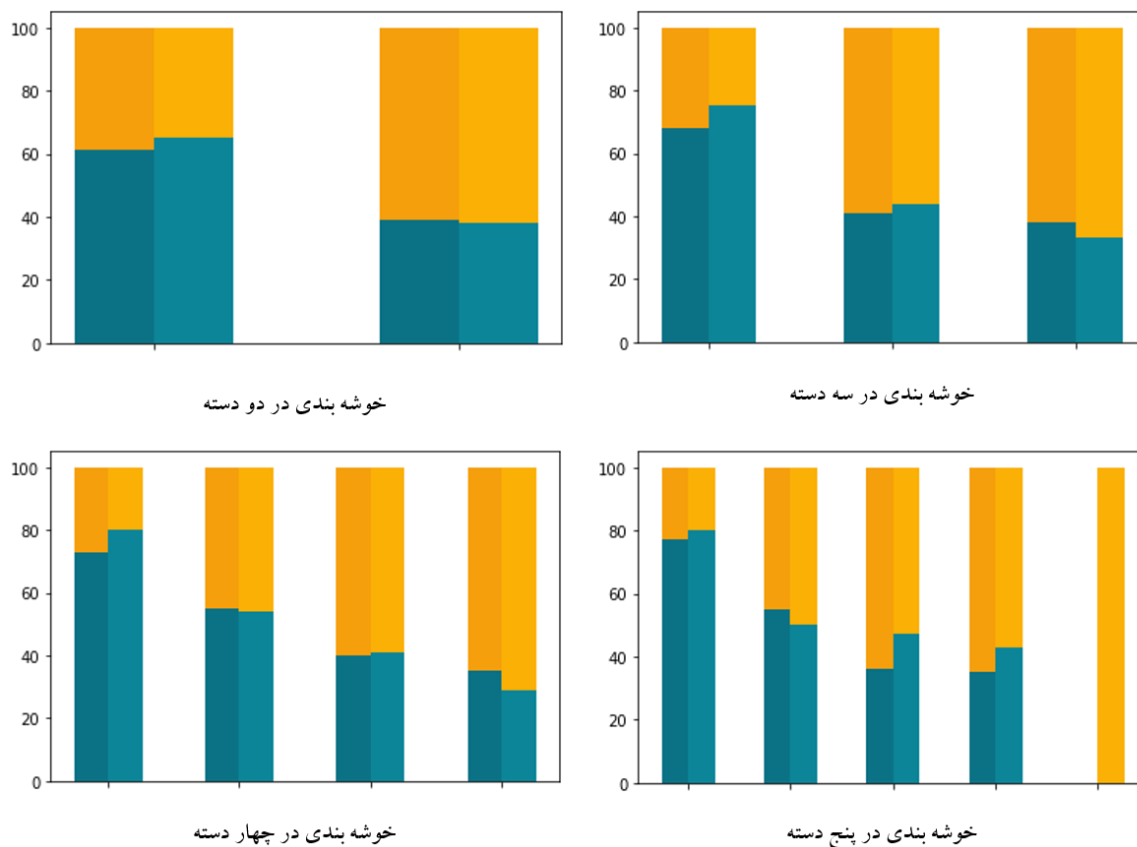
در تصویر ۱-۳ نشان داده شده است که در خوشه ی اول و دوم، بعد از خوشه بندی با ویژگی‌های خوب، درصد افراد سالم افزایش میابد. در خوشه ی چهارم درصد افراد ناسالم کمی افزایش میابد.

با مقایسه ی نتایج بالا میتوان متوجه شد که با افزایش تعداد خوشه‌ها، درصد افراد سالم یا ناسالم یک خوشه افزایش میابد. در بقیه ی خوشه‌ها، که درصد آدم‌های سالم و ناسالم آن بین ۴۰ تا ۶۰ درصد می‌باشد، درصد هیچ کدام از افراد به طور معنا داری تغییر نکرده است.

پس بار دیگر برای مقایسه ی خوشه‌بندی‌ها، خوشه بندی با پنج خوشه انجام می‌شود .

در تصویر ۱-۳ برای خوشه‌بندی در پنج دسته، خوشه‌ای قابل مشاهده است که تمام افراد آن ناسالم اند (خوشه سوم) و خوشه ای که نزدیک به ۸۰ درصد افراد آن سالم می‌باشند.

باید توجه داشت که در نتایجی که خوشه بندی با تمام ویژگی‌ها برای پنج خوشه انجام شد، در یکی از خوشه ها فقط یک گراف مغزی وجود داشت و چون یک گراف معیار مناسبی برای مقایسه نیست، این خوشه قبل از خوشه‌بندی با ویژگی‌های خوب با دیگر خوشه‌ها مقایسه نمی‌شود. میتوان نتیجه گرفت که خوشه‌بندی با پنج خوشه و با تمام ویژگی‌ها نتیجه ی خوبی نمی‌دهد.



راهنمای نمودار: در هر میله رنگ آبی نشاندهنده درصد افراد سالم و رنگ زرد نشان دهنده درصد افراد اوتیسم میباشد. و در هر خوشه میله سمت چپ نتیجه حاصل از تمام ویژگی ها و میله سمت راست نتیجه حاصل از ویژگی هایی است که با آستانه ۰.۰۵ روی p -value های حاصل از تحلیل واریانس انتخاب شده اند.

افراد سالم (تمام ویژگی ها)
افراد اوتیسم (تمام ویژگی ها)
افراد سالم (ویژگی های منتخب)
افراد اوتیسم (ویژگی های منتخب)

تصویر ۱-۳

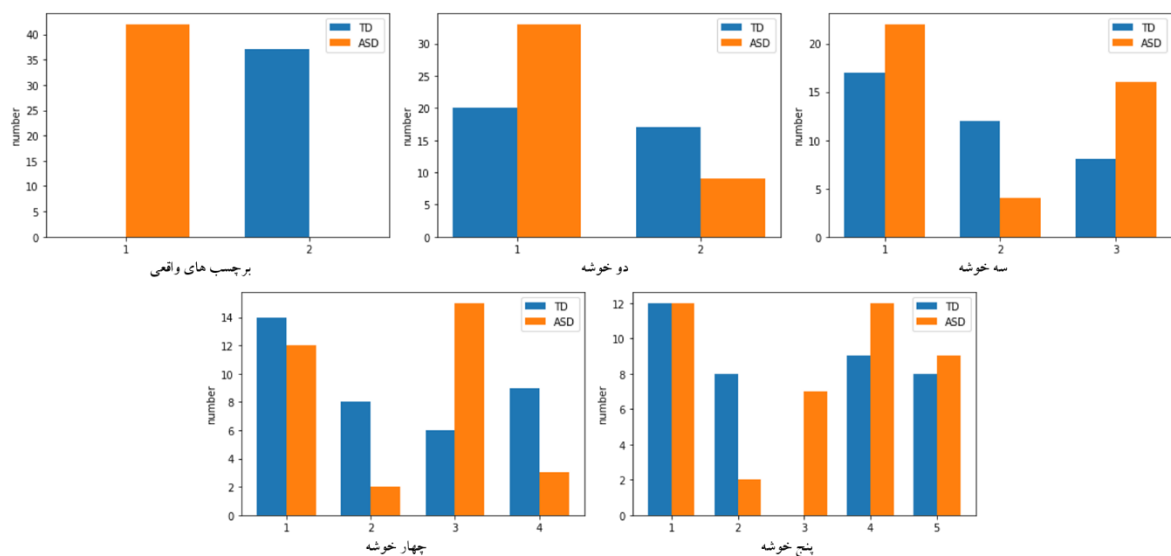
۳-۵- مقایسه ی نتایج حاصل از خوشه ها با برچسب های اصلی

در ادامه قصد بر این است که با بررسی بیشتر روی هر ویژگی و مقادیر آن برای گراف های مغزی موجود در هر خوشه، به دنبال ساختار خاصی بگردیم و بررسی کنیم که آیا ساختار خاصی برای ایجاد تفاوت بین افراد سالم و اوتیسم می باشد یا خیر؟! بدین منظور برای هر ویژگی، نمودار جعبه ای آن رسم شده و آن ها از لحاظ میانه و دامنه تغییرات مقایسه می شوند.

ابتدا پراکندگی افراد سالم و اوتیسم در خوشه ها، از نظر تعداد، نشان داده می شوند و همانطور که در تصویر ۳-۲ مشاهده می شود، در لیبل های اصلی، گروه ۲ افراد سالم هستند و گروه ۱ افراد اوتیسم قرار دارند.

در خوشه بندی در دو دسته، در خوشه ۲، درصد افراد سالم و در خوشه ۱، درصد افراد اوتیسم بیشتر است.

و در سه خوشه بندی دیگر، در خوشه ۲، درصد افراد سالم و در خوشه ۳، درصد افراد اوتیسم بیشتر است.

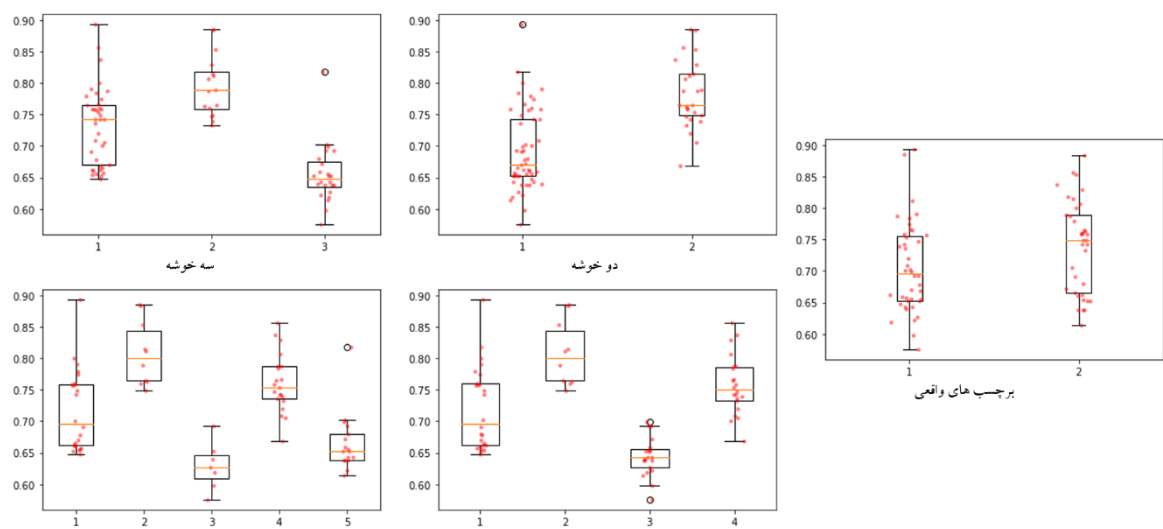


تصویر ۳-۲ - پراکندگی افراد سالم و اوتیسم در خوشه ها و برچسب های اصلی

▪ ضریب خوشگی (ماکسیمم)

به طور تقریبی میتوان گفت که ماکسیمم ضریب خوشگی در افراد سالم نسبت به افراد اوتیسم بیشتر است و میانه‌ی این مقادیر در افراد سالم بالاتر از افراد اوتیسم قرار دارد.

گروه‌هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه‌ی خوشه بندی‌ها) تا حدی نماینده اوتیسم هستند نسبت به گروه‌های ۲ که نماینده افراد سالم در این دسته بندی‌ها هستند، ماکسیمم ضریب خوشگی کمتری دارند.

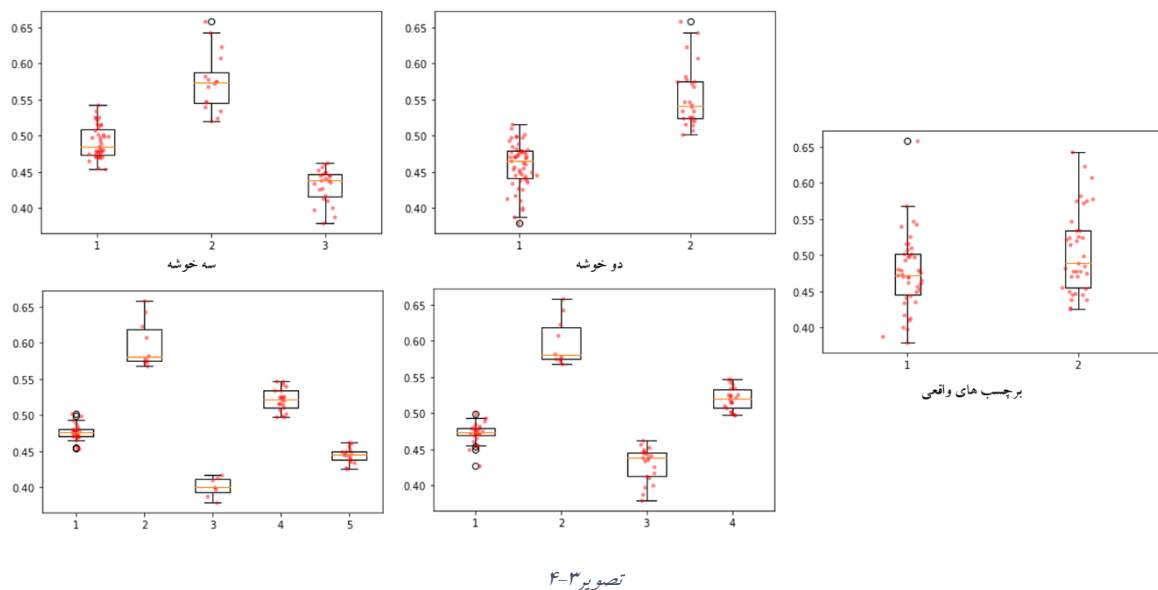


تصویر ۳-۳

▪ ضریب خوشگی (میانہ)

به طور تقریبی میتوان گفت میانہ ضریب خوشگی در افراد سالم نسبت به افراد اوتیسم بیشتر است و با اینکه میانہ ها بسیار به هم نزدیک هستند ولی جعبه ی نمودار در افراد سالم، بالاتر از افراد اوتیسم قرار دارد.

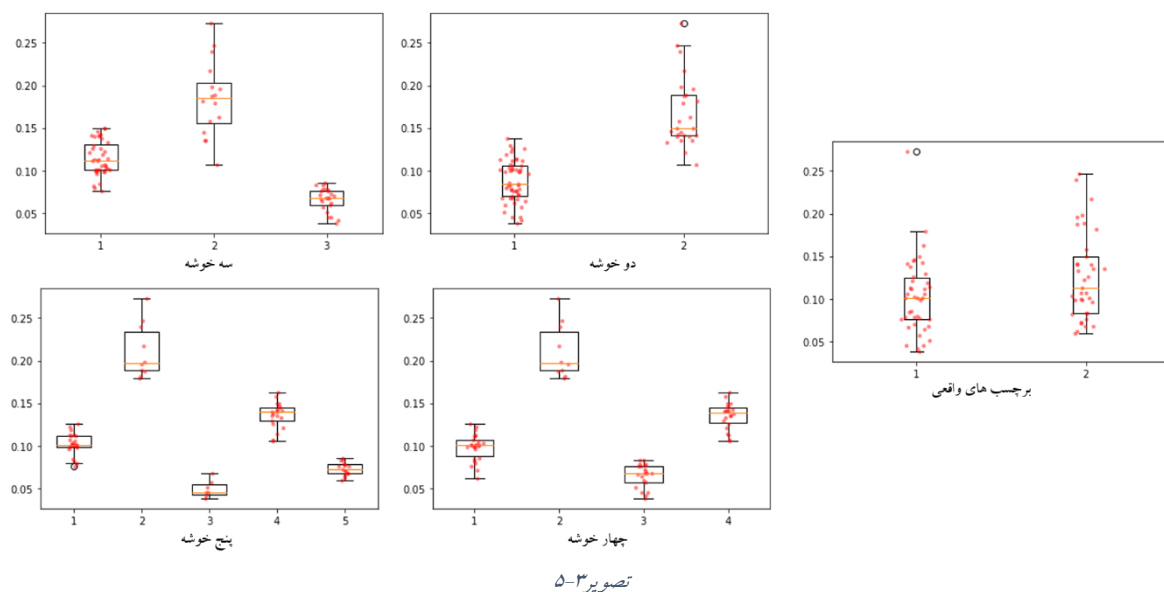
گروه هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند نسبت به گروه های ۲ که نماینده افراد سالم در این دسته بندی ما هستند، میانہ ضریب خوشگی کمتری دارند.



▪ مثلث ها (ماکسیمم)

به طور تقریبی میتوان گفت ماکسیمم مثلث ها در افراد سالم نسبت به افراد اوتیسم بیشتر است و با اینکه میانہ ها بسیار به هم نزدیک هستند ولی جعبه ی نمودار در افراد سالم، بالاتر از افراد اوتیسم قرار دارد.

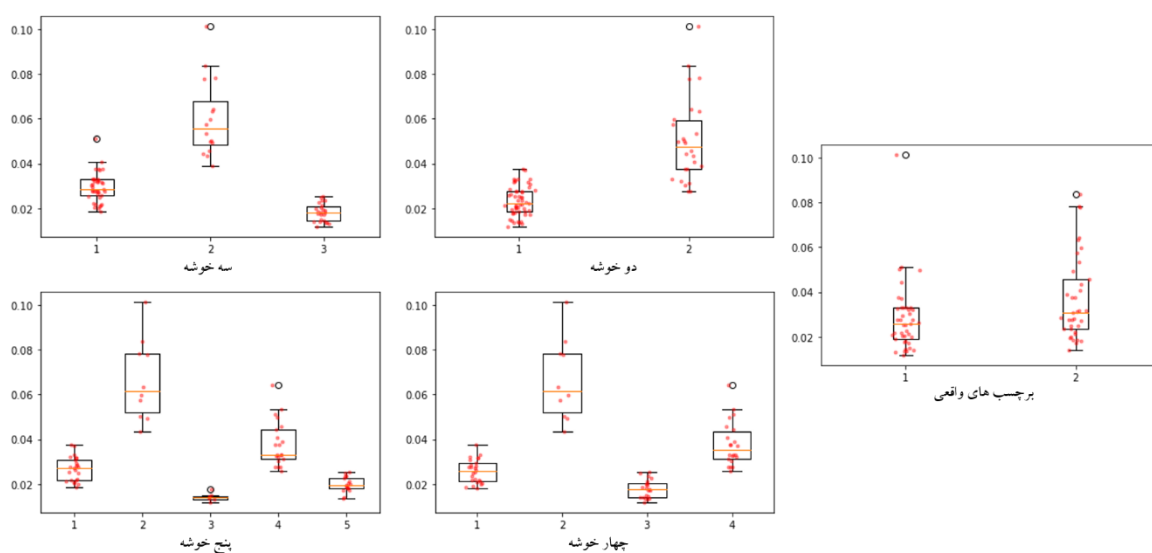
گروه هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند نسبت به گروه های ۲ که نماینده افراد سالم در این دسته بندی ما هستند، ماکسیمم مثلث ها کمتری دارند.



▪ مثلث ها (میانه)

به طور تقریبی میتوان گفت میانه تعداد مثلث ها در افراد سالم نسبت به افراد اوتیسم کمی بیشتر است ولی طیف پراکندگی سالم بیشتر از اوتیسم میباشد ولی میانه هایی بسیار نزدیک به هم دارند.

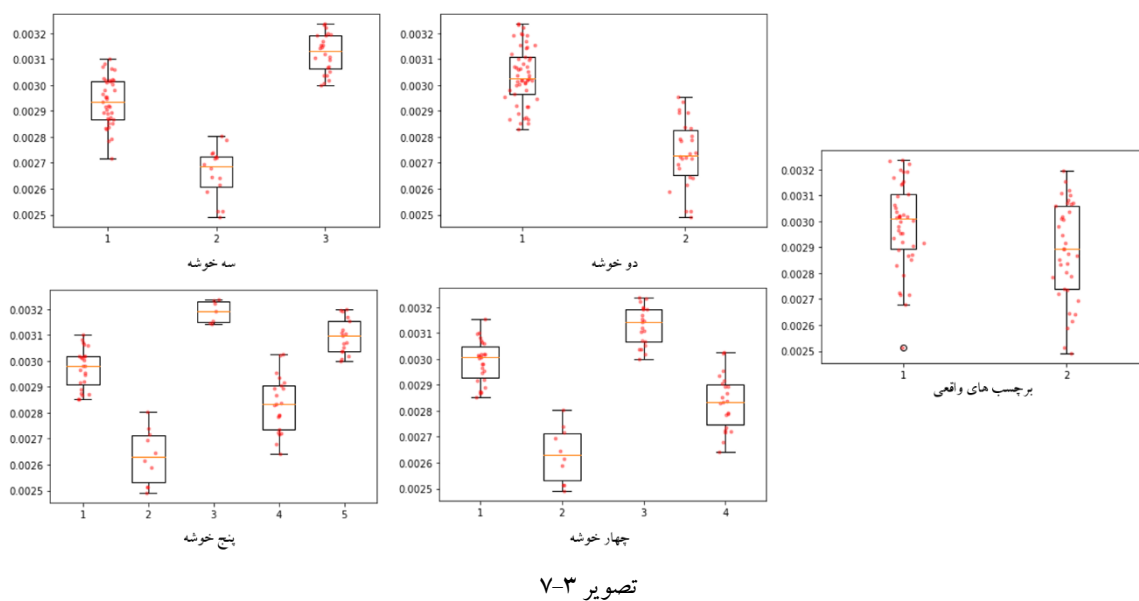
گروه هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند نسبت به گروه های ۲ که نماینده افراد سالم در این دسته بندی ما هستند، میانه مثلث های کمتری دارند.



▪ مرکزیت بینابینی (میانه)

به طور تقریبی میتوان گفت میانه مرکزیت بینابینی در افراد سالم به نسبت افراد اوتیسم کمی کمتر است و میانه ی مقادیر هر خوشه در افراد سالم پایین تر از افراد اوتیسم میباشد ولی طیف پراکندگی سالم بیشتر از اوتیسم میباشد.

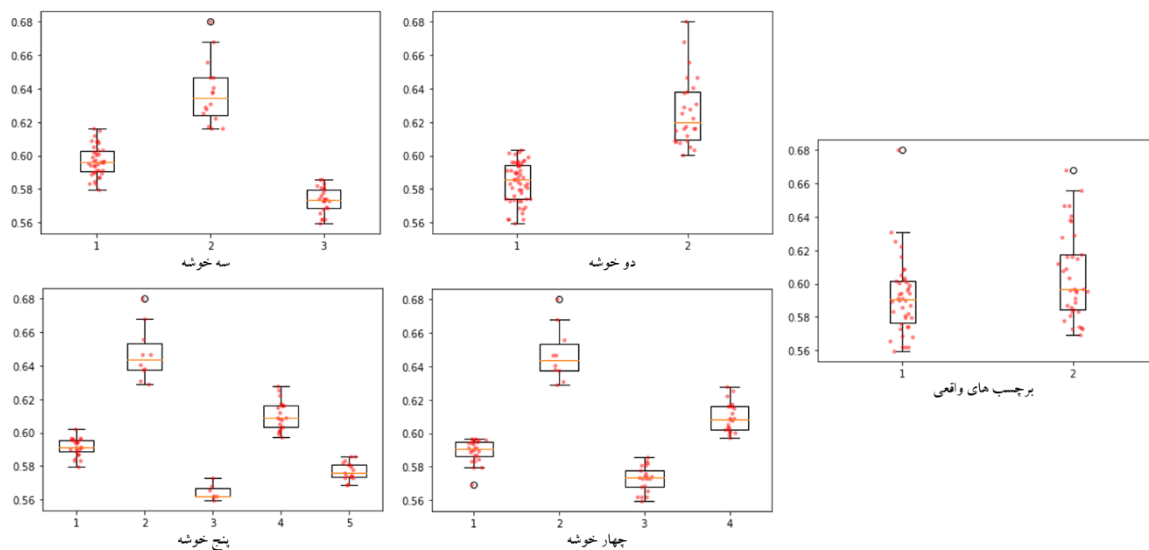
گروه هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند نسبت به گروه های ۲ که نماینده افراد سالم در این دسته بندی ما هستند، میانه ی مرکزیت بینابینی بیشتری کمتری دارند. (یا جعبه ی آنها بالا تر است و یا پراکندگی آنها کمتر است).



▪ مرکزیت بر مبنای نزدیکی (ماکسیمم)

به طور تقریبی میتوان گفت ماکسیمم مرکزیت بر مبنای نزدیکی در افراد سالم نسبت به افراد اوتیسم کمی بیشتر است ولی میانه های بسیار نزدیک به هم دارند.

گروه هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند نسبت به گروه های ۲ که نماینده افراد سالم در این خوشه بندی هستند، ماکسیمم مرکزیت بر مبنای نزدیکی کمتری دارند.

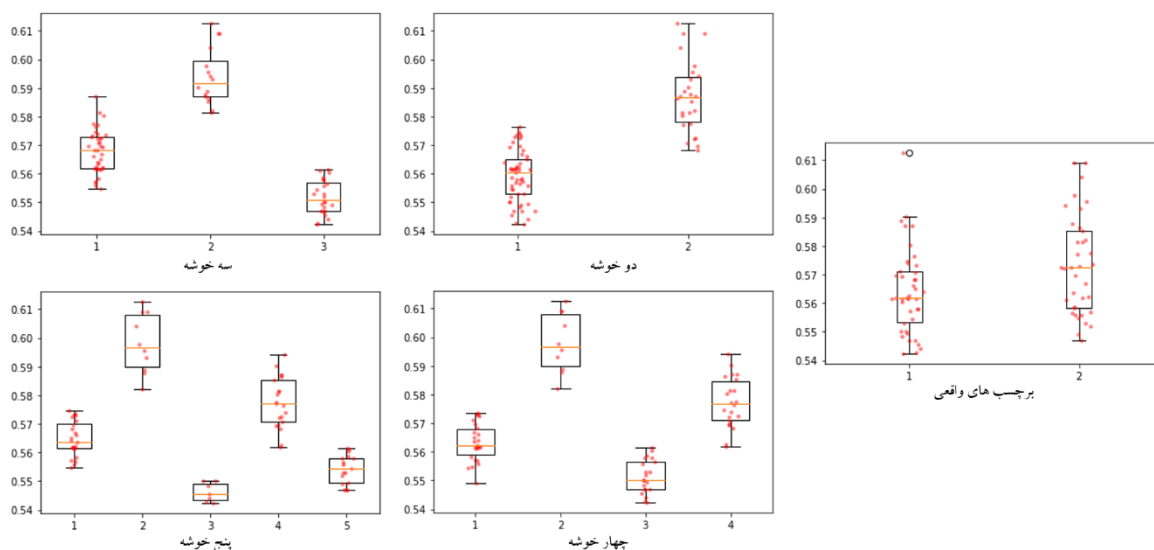


تصویر ۸-۳

▪ مرکزیت بر مبنای نزدیکی (میان)

به طور تقریبی گفت میانه مرکزیت بر مبنای نزدیکی در افراد سالم نسبت به افراد اوتیسم بیشتر است.

گروه هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند نسبت به گروه های ۲ که نماینده افراد سالم در این دسته بندی ما هستند، میانه مرکزیت بر مبنای نزدیکی کمتری دارند.

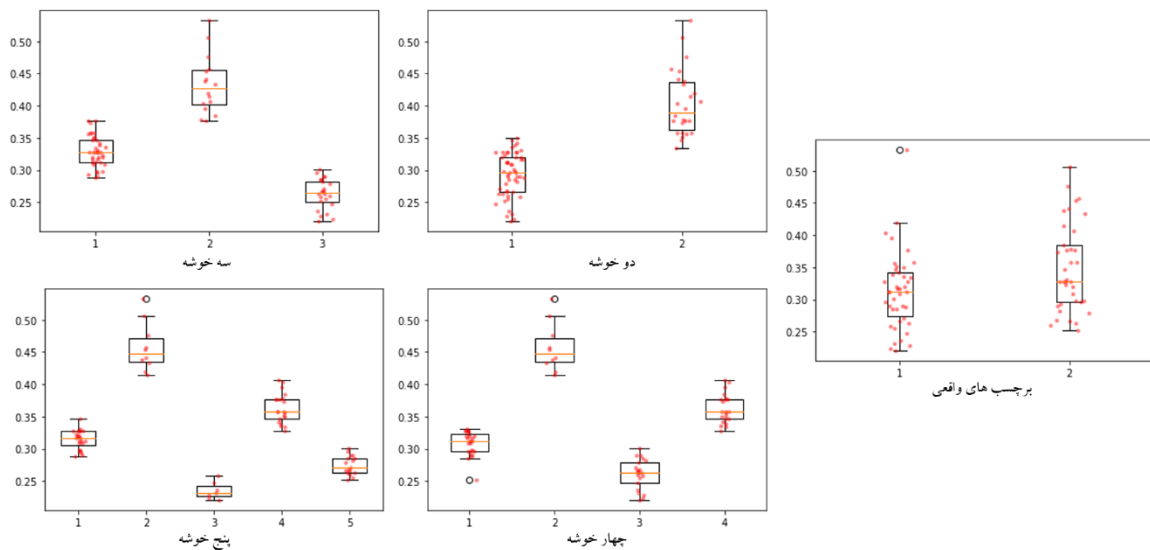


تصویر ۹-۳

▪ مرکزیت بر مبنای درجه (ماکسیمم)

به طور تقریبی میتوان گفت ماکسیمم مرکزیت بر مبنای درجه در افراد سالم نسبت به افراد اوتیسم بیشتر است.

گروه‌هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند، نسبت به گروه های ۲ که نماینده افراد سالم در این خوشه‌بندی هستند، ماکسیمم مرکزیت بر مبنای درجه کمتری دارند.

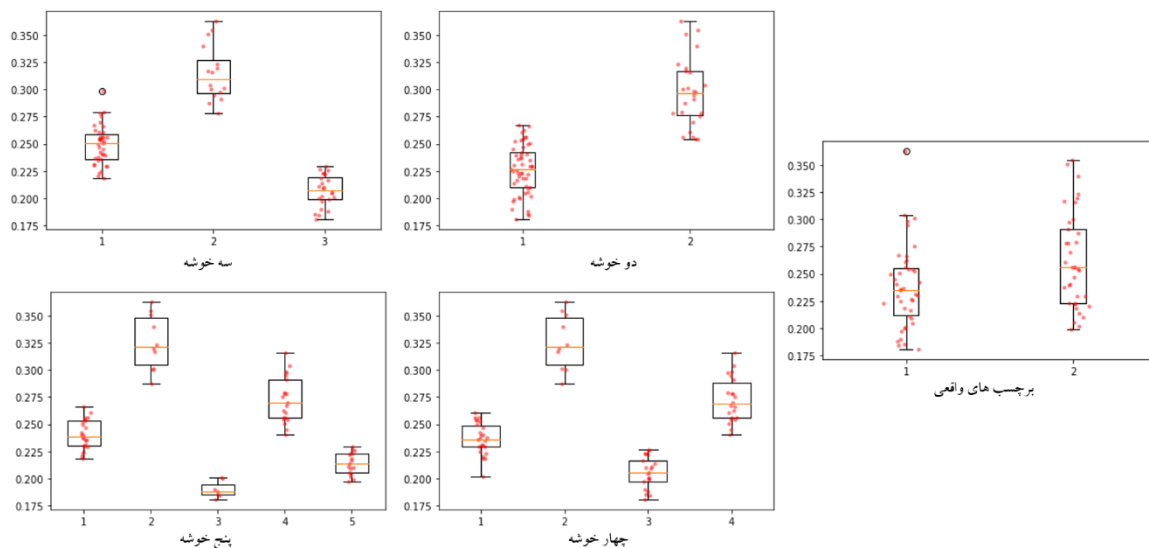


تصویر ۳-۱۰

▪ مرکزیت بر مبنای درجه (میانه)

به طور تقریبی میتوان گفت میانه مرکزیت بر مبنای درجه در افراد سالم نسبت به افراد اوتیسم بیشتر است.

گروه‌هایی که (با برچسب ۱ در دو خوشه و با برچسب ۳ در بقیه ی خوشه بندی ها) تا حدی نماینده اوتیسم هستند، نسبت به گروه های ۲ که نماینده افراد سالم در این دسته بندی ما هستند، میانه مرکزیت بر مبنای درجه کمتری دارند.

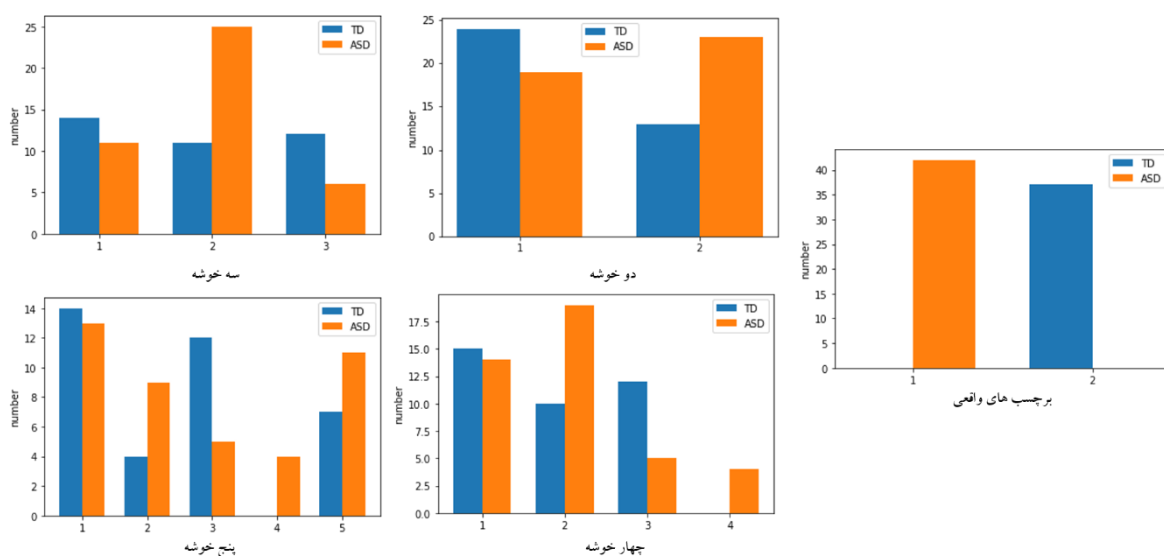


تصویر ۱۱-۳

۳-۵-۱- مقایسه ی نتایج حاصل از خوشه ها با آستانه ای متفاوت

مراحل و بررسی های قبلی بار دیگر با یک آستانه ی متفاوت (۰.۳۵ تا ۰.۴) انجام شد که به طور نمونه یکی از نتایج آن به شرح زیر است.

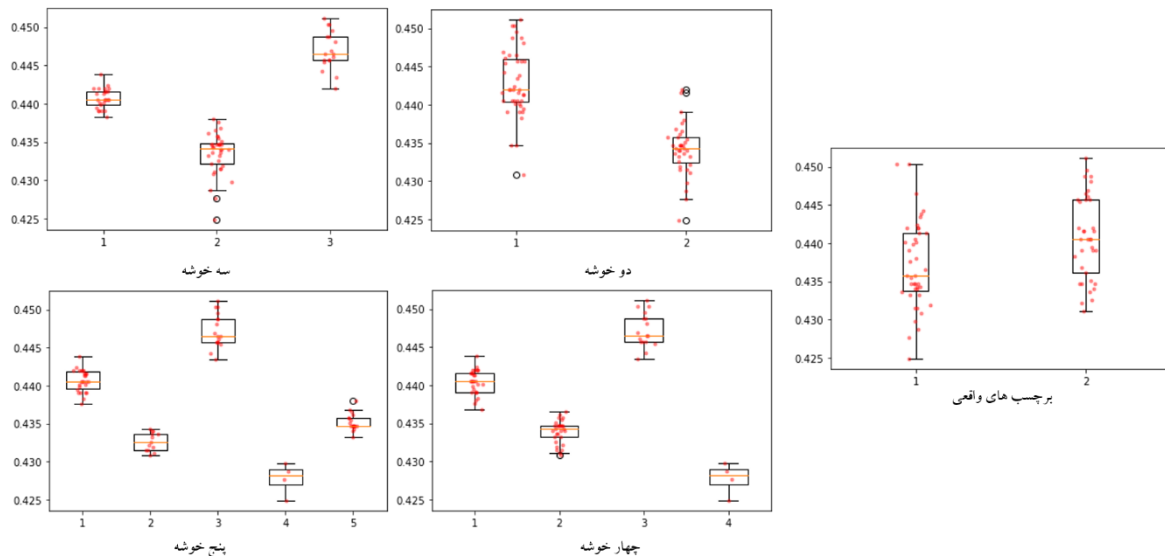
- در برچسب های اصلی، گروه ۲ افراد سالم هستند و گروه ۱ افراد اوتیسم قرار دارند.
- در خوشه بندی در دودسته، در خوشه ی ۲، درصد افراد اوتیسم و در خوشه ی ۱، درصد افراد سالم بیشتر است.
- در خوشه بندی در سه دسته، در خوشه ی ۲، درصد افراد اوتیسم و در خوشه ی ۳، درصد افراد سالم بیشتر است.
- و در دو خوشه بندی دیگر، در خوشه ی ۳، درصد افراد سالم و در خوشه ی ۴، درصد افراد اوتیسم بیشتر است.



تصویر ۱۲-۳

▪ مرکزیت بر مبنای نزدیکی (میان)

به طور تقریبی میتوان گفت میانه مرکزیت بر مبنای نزدیکی در افراد سالم نسبت به افراد اوتیسم بیشتر است. گروه هایی که تا حدی نماینده اوتیسم هستند، نسبت به گروه هایی که نماینده افراد سالم در این خوشه بندی هستند، میانه مرکزیت بر مبنای نزدیکی کمتری دارند.



تصویر ۳-۱۳

۳-۵-۲- نتیجه گیری

در تمام ویژگی های ذکر شده در این قسمت، باید در نظر داشت که برای برچسب های اصلی، هیچ یک از جعبه ها کاملاً از یک دیگر جدا نشده اند ولی میتوان دید که به طور تقریبی مقادیر ویژگی ها در افراد سالم (به جز میانه ی مرکزیت بینایی) نسبت به اوتیسم بیشتر بوده است (طیف تغییرات مقادیر بیشتر بوده یا جعبه در مکان بالاتری قرار میگیرد و یا میانه بیشتری داشته است). در میانه ی مرکزیت بینایی، جعبه افراد اوتیسم بالاتر از افراد سالم قرار دارد که بدان معنی است که مقادیر این ویژگی در افراد اوتیسم به نسبت افراد سالم بیشتر است.

برای خوشه بندی های صورت گرفته این نتایج مشاهده میشود:

در صورتی که اختلاف دو خوشه مقدار قابل توجهی باشد (به طور مثال خوشه ۲ و ۳ در پنج خوشه ی میانه مرکزیت بر مبنای درجه) در این صورت از روند لیل های اصلی پیروی میکند یعنی اگر در لیل های اصلی جعبه سالم بالاتر از اوتیسم قرار دارد در این حالت نیز خوشه ای که تعداد افراد سالم به اوتیسم آن بیشتر است نیز، بالاتر قرار میگیرد.

اما اگر در خوشه ای اختلاف افراد سالم و ناسالم تا حدی نزدیک به یک دیگر باشد، نمیتوان به طور دقیق اظهار نظر کرد.

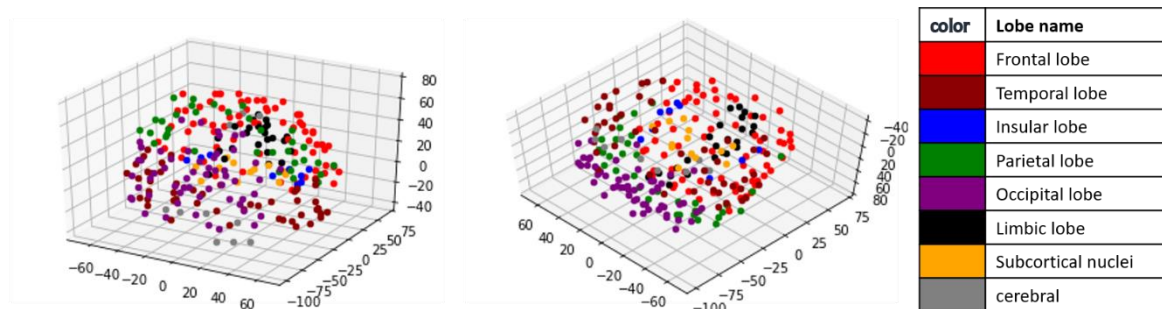
نتیجه گیری نهایی:

- ۱- به طور کلی میانه مثلث‌های روی هر راس در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.
- ۲- به طور کلی ماکسیمم ضریب خوشگی در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.
- ۳- به طور کلی میانه مرکزیت بر مبنای درجه در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.
- ۴- به طور کلی میانه مرکزیت بینایی در افراد سالم کمتر از افراد مبتلا به اوتیسم می باشد.
- ۵- به طور کلی میانه مرکزیت بر مبنای نزدیکی در افراد سالم بیشتر از افراد مبتلا به اوتیسم می باشد.

۳-۶- بررسی منطقه ای راس های منتخب

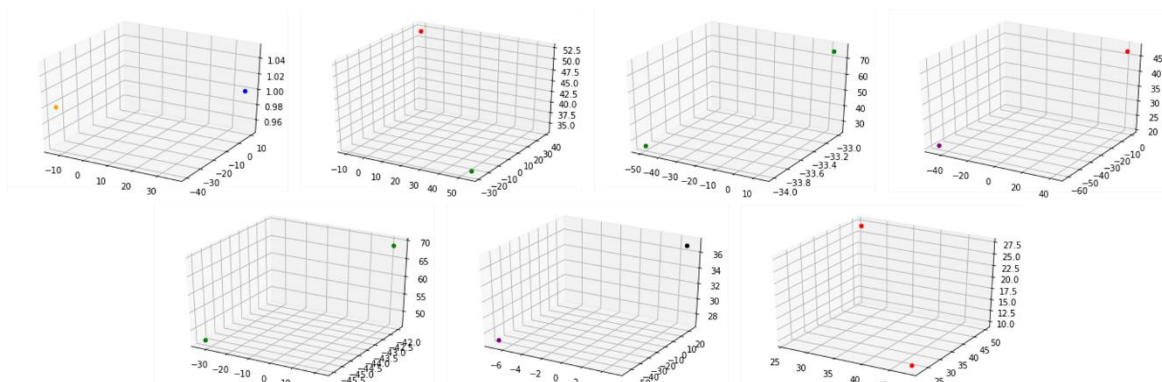
بررسی دیگری در این زمینه انجام شد با هدف اینکه آیا رئوس خاصی از گراف (به طور مثال رئوسی از گراف که در آن‌ها مقدار میانه ی مرکزیت بینایی اتفاق می افتد) در قطعه‌ی خاصی از مغز قرار دارد یا خیر؟!.

در تصویر ۳-۱۴، رئوس مربوط به گراف و مناطق مغزی آن‌ها را مشاهده میکنید.



تصویر ۳-۱۴ - راس ها و مناطق مغزی آنها

ابتدا رئوسی که در آنها میانه ی ضریب خوشگی اتفاق افتاده است را پیدا کرده و با رنگ های متفاوت نشان می‌دهیم که هر یک از آنها در چه منطقه ای از مغز قرار دارد. نتایج حاصل از این بررسی برای یکی از خوشه ها، که تمام افراد آن اوتیسم می‌باشد، به شرح زیر است :



تصویر ۳-۱۵ - راس ها و مناطق مغزی آنها

همانطور که در تصویر قابل مشاهده است، در همه ی افراد این خوشه، منطقه ی مشترکی (رنگ مشترکی) وجود ندارد. پس میتوان نتیجه گرفت که این رئوس، لزوما در مناطق خاصی از گراف قرار ندارند.

فصل چهارم

مقایسه و جمع بندی نهایی

۴-۱- مقایسه

در پژوهش دیگری در همین زمینه همزمان با انجام این پروژه صورت گرفت که از الگوریتم خوشه بندی متفاوتی استفاده شده است که مبتنی بر خوشه بندی سلسه مراتبی با روش *agglomerative* است. در ادامه ترکیب و مقایسه نتایج هر دو روش خوشه بندی آورده خواهد شد.

به منظور مقایسه دو روش در ابتدا باید دانست که هر دو روش روی مجموعه داده ای موجود تا حدی دقت برابری را نتیجه می دهند لذا نمیتوان به طور دقیق گفت کدام روش ها بهتر یا بدتر عمل میکنند. اما با توجه به نتایج به دست آمده با این دیتاست، در یکی از خوشه های مدل ترکیبی گوسی توانسته به اختلاف ۱۰۰ درصدی برای اوتیسم برسد که بدین معناست که تمام افراد حاضر در یکی از گروه های خوشه بندی ها تمام افراد اوتیسم هستند. و بهترین درصد هم برای خوشه بندی افراد سالم با استفاده از این روش ۸۰٪ می باشد. در روش خوشه بندی سلسله مراتبی اعداد بالا به ترتیب ۷۰٪ و ۸۰٪ می باشد که از این حیث تا حدی الگوریتم ترکیبی گوسی بهتر عمل کرده است.

حال از این دو روش چه استفاده هایی می توان کرد. یکی از راهکار ها هنگام تشخیص برچسب کانکتوم های مغزی افراد میباشد بدین صورت که داده جدید به همراه داده های قبلی مجدداً خوشه بندی میشود با استفاده از هر دو روش، اگر داده جدید در هر دو روش به دسته ای تعلق گرفت که آن دسته به گونه ای نماینده ی افراد اوتیسم باشد (و برعکس) با احتمال خوبی می توان گفت که داده جدید نیز احتمالاً متعلق به فرد مبتلا به اوتیسم است (و برعکس)

۴-۲- جمع بندی نهایی

هدف اصلی این پروژه مقایسه ی این روش با کلاسیفیکیشن و بررسی میزان عملکرد خوشه بندی بوده است.

از نتایج ذکر شده در پروژه میتوان پی برد که به دلیل گسترده بودن طیف بیماران اوتیسم، ویژگی های استخراج شده از گراف مغزی افراد ناسالم با افراد سالم اشتراکات زیادی دارند و نمیتوان آن ها را به طور کامل از یکدیگر تفکیک نمود.

با این حال در برخی از ویژگی ها افراد سالم و ناسالم تفاوت محسوسی دارند که قابل تشخیص است و میتواند بخشی از افراد سالم را از ناسالم تفکیک نمود.

همچنین از نتایج این پژوهش این است که در ویژگی مرکزیت بینایی، به طور میانگین و در مقادیر میانه ی آن (یعنی در گراف مغزی آن، مقداری که در میانه ی این ویژگی قرار میگیرند) در افراد اوتیسم بیشتر از افراد سالم است و در ویژگی مرکزیت بر مبنای نزدیکی و بر مبنای درجه، به طور میانگین مقادیر و در مقادیر میانه ی آن در افراد اوتیسم کمتر از افراد سالم است.

همچنین از نتایج این پژوهش این است که به طور کلی مرکزیت بینایی در افراد سالم کمتر از افراد مبتلا به اوتیسم میباشد. که به این معنا میباشد که بطور میانگین مسیر های بیشتری از رئوس مربوط به گراف های مغزی افراد اوتیسم میگذرد. و به طور کلی ضریب خوشگی در افراد سالم بیشتر از افراد مبتلا به اوتیسم میباشد که به این معنا می باشد که بطور میانگین رئوس گراف های مغزی افراد سالم تمایل بیشتری به ایجاد یک خوشه دارند و در ویژگی مرکزیت بر مبنای نزدیکی و بر مبنای درجه، به طور کلی مقادیر آن ها در افراد اوتیسم کمتر از افراد سالم است.

و در نهایت با بررسی مناطق رئوس منتخب میتوان نتیجه گرفت که این رئوس لزوماً مربوط به مناطق خاصی از گراف نمی باشند.

فصل پنجم

ضمیمه ها

در این قسمت جزئیات برخی از نتایج آورده شده است.

۵-۱- جداول

این جدول در صفحه ی بعد قابل مشاهده است.

راهنمای جدول :

- مقادیر نشان داده شده با رنگ قرمز *** به دلیل اینکه مقادیری هستند که در همه ی کانکتوم ها مقادیر یکسانی دارند، معیار مناسبی برای مقایسه نمیباشند.
- مقادیر نشان داده شده با رنگ سبز کمرنگ *** نشان دهنده ی مناسب بودن آن ویژگی در آن دسته بندی میباشد.
- معیار های نشان داده شده با رنگ سبز *** نشان دهنده ی معیار هایی هستند که هم در کلاسترینگ ها و هم برای لیبیل های اصلی، جدا کننده های نسبتاً خوبی میباشند
- معیار های نشان داده شده با رنگ نارنجی *** نشان دهنده ی معیار هایی هستند که در کلاسترینگ ها جدا کننده ی خوبی هستند ولی هم برای لیبیل های اصلی، جدا کننده های خوبی نمیباشند.

Features			2 clusters	3 clusters	4 clusters	Real labels
Node strength	0	Max	6.9302e-06	3.7011e-10	3.2639e-10	0.94125648
	1	Chart3	6.5912e-06	4.2010e-07	6.7147e-06	0.62668749
	2	Med	0.77209863	0.38793181	0.72474801	0.87282779
	3	Mean	0.65610784	0.33473006	0.12056660	0.61529760
	4	Chart1	2.2342e-05	3.7817e-07	2.9397e-07	0.78018553
	5	Min	9.6725e-07	1.1367e-12	1.7258e-10	0.93273139
Eccentricity	6	Max	0.65901993	0.31144037	0.44570043	0.37553030
	7	Chart3	nan	nan	nan	nan
	8	Med	0.00042758	0.02032502	0.04908686	0.06363024
	9	Mean	9.3559e-08	2.6818e-05	5.2075e-05	0.09677209
	10	Chart1	0.00015626	0.00415587	0.00562021	0.64694607
	11	Min	nan	nan	nan	nan
Triangles	12	Max	5.8428e-11	3.5824e-14	1.4461e-25	0.02849925
	13	Chart3	5.3843e-11	2.5954e-17	5.4321e-26	0.02794224
	14	Med	1.0158e-11	5.5011e-17	1.5466e-18	0.02777244
	15	Mean	3.6767e-12	5.4973e-18	3.9781e-25	0.02618098
	16	Chart1	5.7697e-11	4.9902e-11	4.1107e-07	0.00879690
	17	Min	0.18381211	0.34142303	0.02598061	0.82543615
Clustering	18	Max	0.00020321	2.4254e-05	6.3371e-09	0.03906564
	19	Chart3	1.7359e-10	2.7442e-14	1.1882e-24	0.03664204
	20	Med	4.5937e-11	2.5221e-15	3.6337e-25	0.01836195
	21	Mean	4.3374e-12	2.9899e-16	2.8585e-25	0.01678299
	22	Chart1	6.3767e-12	1.5440e-15	5.1937e-19	0.00853054
	23	Min	0.31523237	0.05324934	0.03038160	0.16504208
Betweenness centrality	24	Max	0.00077406	4.7975e-05	0.00024072	0.11308404
	25	Chart3	9.8348e-13	3.9600e-13	3.6934e-13	0.03121171
	26	Med	4.4255e-14	9.4242e-12	5.9235e-13	0.03081895
	27	Mean	2.2377e-15	8.1552e-17	2.2771e-19	0.01545228
	28	Chart1	6.4447e-06	9.5050e-07	7.1972e-11	0.04962542
	29	Min	0.35683181	0.06787247	0.08268949	0.09328671
Closeness centrality	30	Max	6.1065e-11	1.1765e-11	1.9962e-23	0.01252525
	31	Chart3	7.5565e-13	1.0263e-17	3.5415e-25	0.01732183
	32	Med	4.3509e-13	5.1796e-16	4.9889e-16	0.01536507
	33	Mean	3.9442e-15	1.0924e-17	4.9468e-21	0.01553695
	34	Chart1	7.3200e-11	1.4599e-08	2.7783e-06	0.04955433
	35	Min	0.25028078	0.09723020	0.04839048	0.79275168
Degree centrality	36	Max	1.2674e-10	4.6536e-11	4.2559e-23	0.01014607
	37	Chart3	1.5851e-12	2.2197e-17	2.4895e-25	0.01466164
	38	Med	4.4024e-13	2.6616e-16	5.2269e-17	0.01334291
	39	Mean	1.6596e-14	7.4977e-18	3.2903e-22	0.01253730
	40	Chart1	2.8751e-11	2.5691e-09	1.0084e-06	0.01541044
	41	Min	0.48306701	0.13232514	0.01261812	0.77442518
Modularity	42		0.41368618	0.95445428	0.80101064	0.87975531
Global efficiency	43		6.3108e-15	1.3704e-17	2.6981e-21	0.01327290

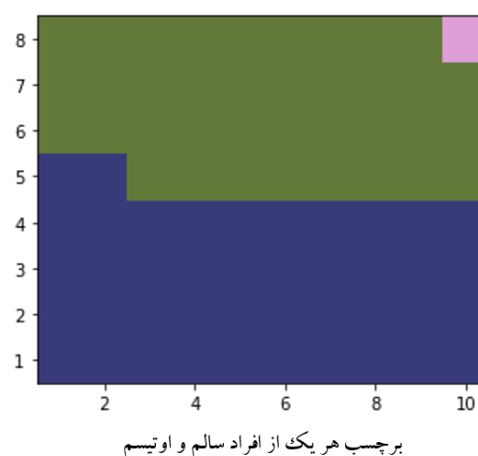
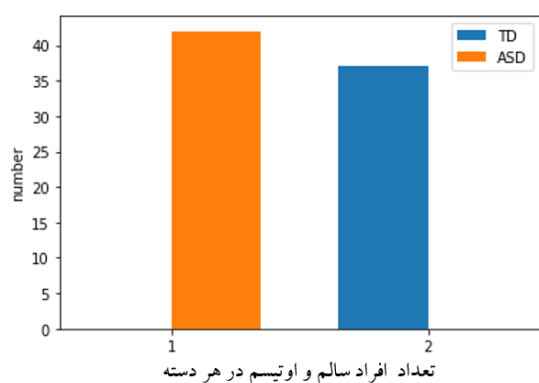
۵-۲- تصاویر

۵-۲-۱- مقایسه خوشه‌ها (قبل و بعد از خوشه بندی با ویژگی های منتخب)

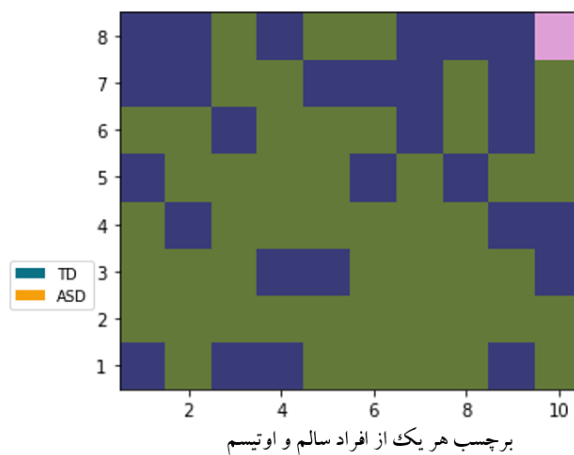
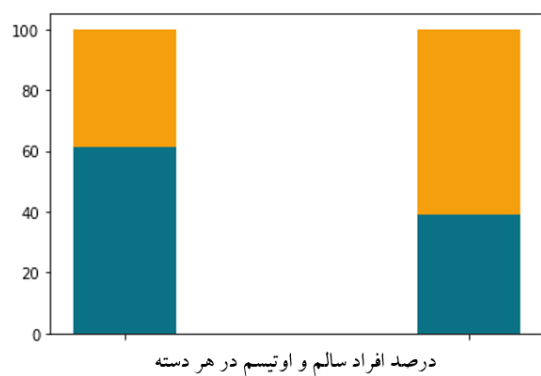
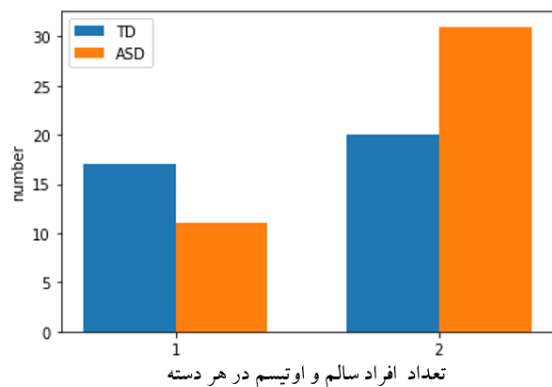
راهنما: در هر یک از تصاویر زیر دو نمودار میله ای مشاهده می‌شود که در نمودار بالا سمت چپ، نشان می‌دهد که در هر دسته چه تعداد فرد سالم و چه تعداد فرد اوتیسم می‌باشد و نمودار پایین سمت چپ نشان می‌دهد که در هر دسته چند درصد از افراد سالم و چند درصد اوتیسم می‌باشد.

افراد سالم با رنگ آبی و افراد اوتیسم با رنگ نارنجی نمایش داده شده اند.

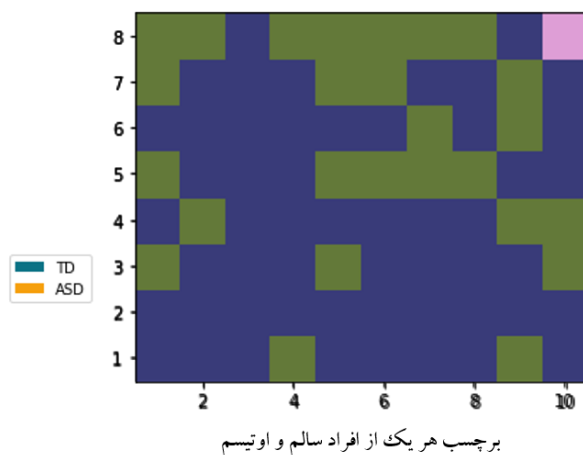
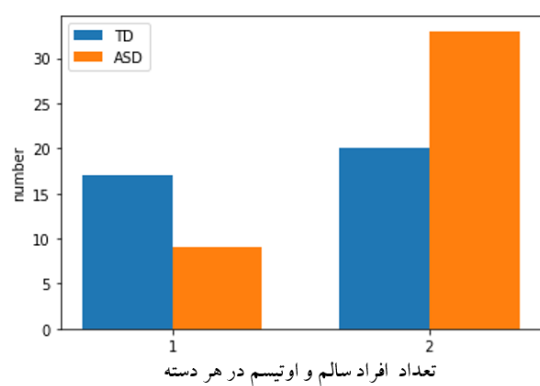
در نمودار سمت راست هم برچسب هر یک از افراد با رنگ های مختلف نشان داده شده است.



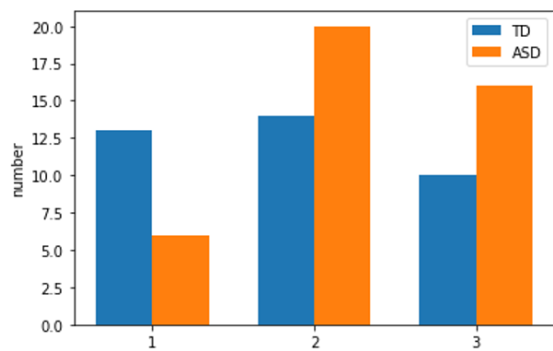
تصویر ۵-۱ - برچسب های اصلی



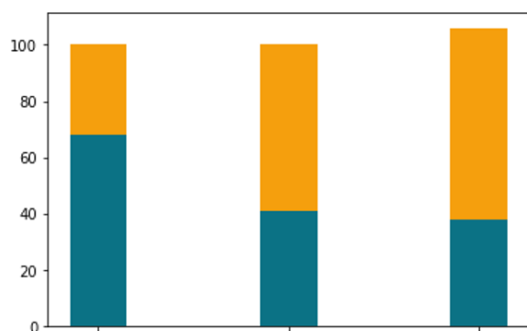
تصویر ۵-۲ - خوشه بندی در دو دسته با استفاده از تمام ویژگی ها



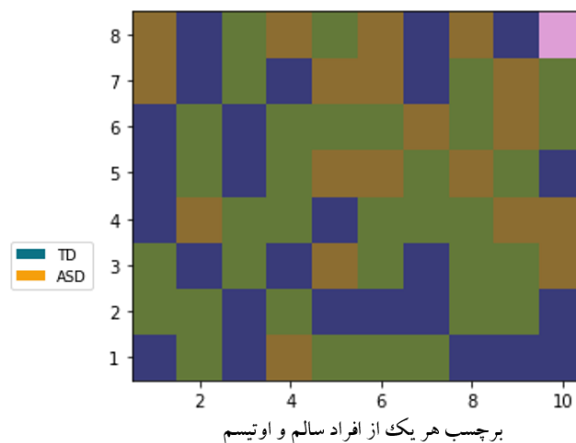
تصویر ۵-۳ - خوشه بندی در دو دسته با استفاده از ویژگی های منتخب



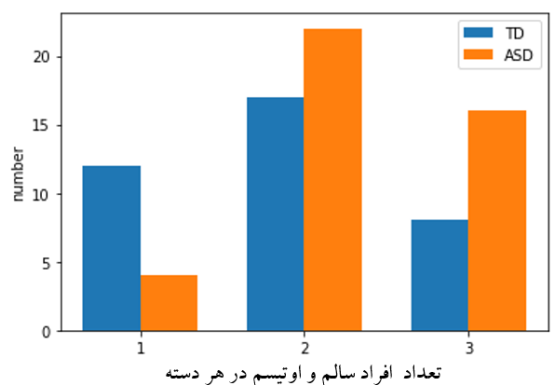
تعداد افراد سالم و اوتیسم در هر دسته



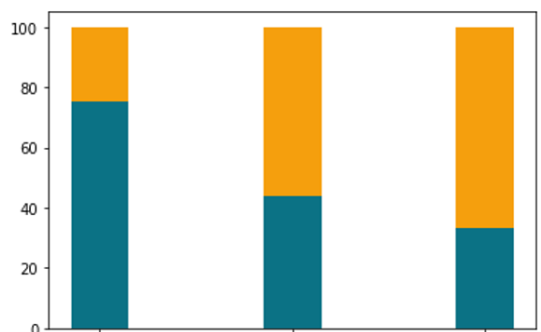
درصد افراد سالم و اوتیسم در هر دسته



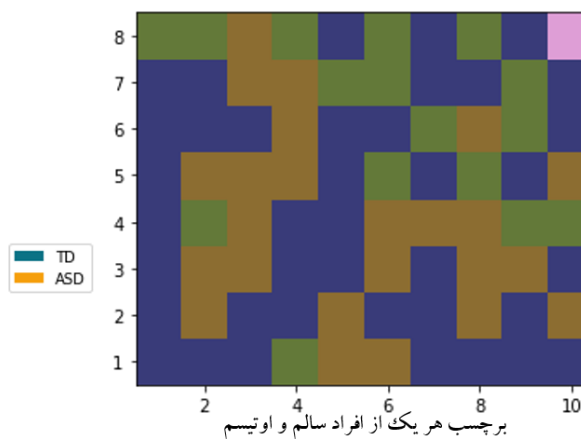
تصویر ۴-۵ - خوشه بندی در سه دسته با استفاده از تمام ویژگی ها



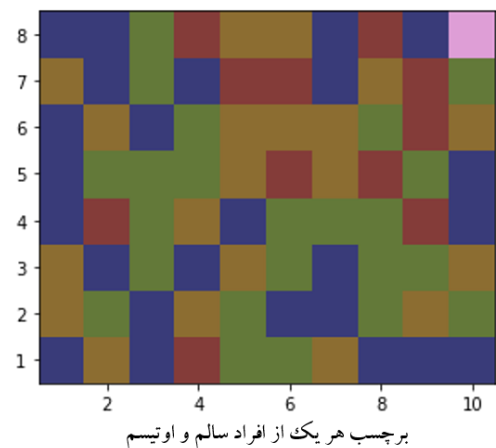
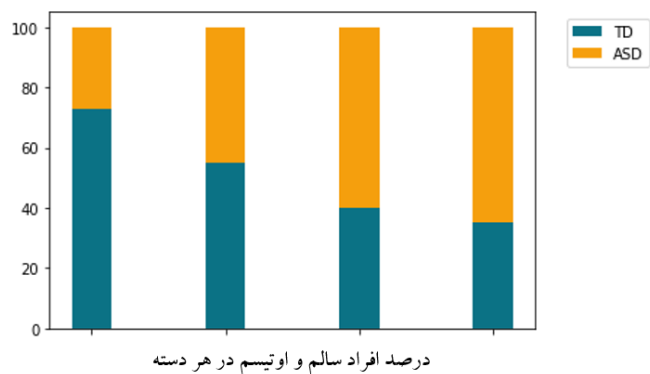
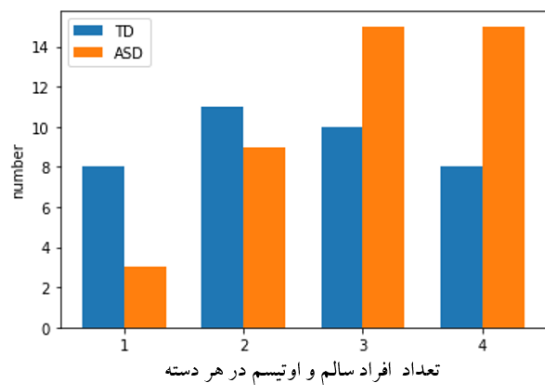
تعداد افراد سالم و اوتیسم در هر دسته



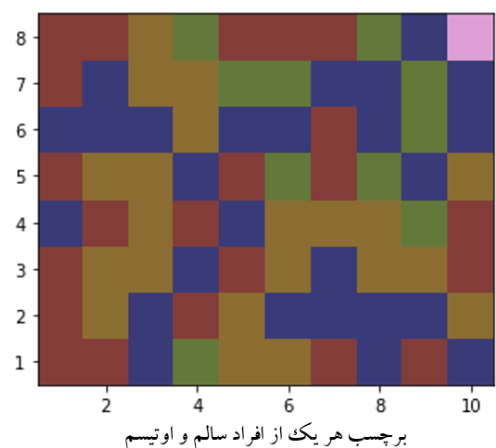
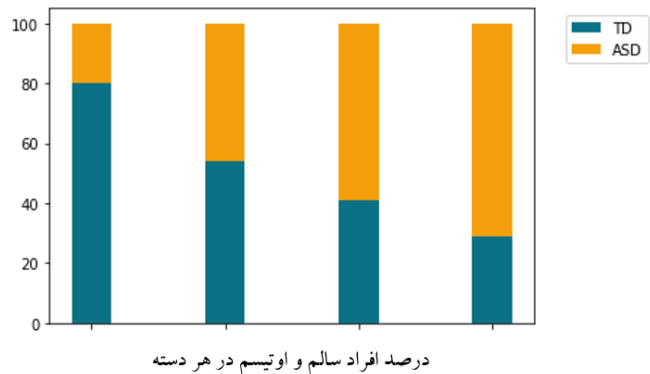
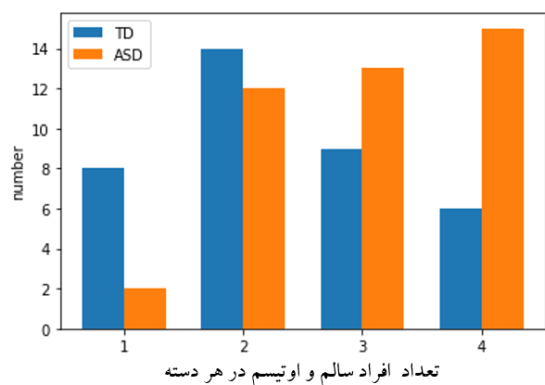
درصد افراد سالم و اوتیسم در هر دسته



تصویر ۵-۵ - خوشه بندی در سه دسته با استفاده از ویژگی های منتخب



تصویر ۵-۶ - خوشه بندی در سه دسته با استفاده از تمام ویژگی ها



تصویر ۵-۷ - خوشه بندی در چهار دسته با استفاده از ویژگی های منتخب

فصل ششم

مراجع

- [1] *Analysis of Variance* (G. Larson, Jan 2008)
- [2] *strategies for detection of autism syndrome disorder*
- [3] *Complex Brain Network Analysis and Its Applications to Brain Disorders*
- [4] *Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention- Deficit/Hyperactivity Disorder Using Functional and Structural MRI*
- [5] *A Physarum Centrality Measure of the Human Brain Network*
- [6] *Diagnosis of autism spectrum disorder based on complex network features. Computer Methods and Programs in Biomedicine*
- [7] *Analysis of brain signals based on complex networks to improve the diagnosis of autism*
- [8] *Documentation* <https://scikit-learn.org/stable/modules/mixture.html>
- [9] *Documentation* <https://networkx.org/documentation/stable/reference/algorithms/>