

To propose a new method for predicting Alzheimer disease in MCI subjects based on the selected features via sparse coding

Paper presenter: Morteza Tourandaz¹

Morteza Tourandaz¹, Muhammad Yousefnezhad², Saber Nourian³

1- Department of Computer Science and Artificial Intelligence, Rouzbahan University, Sari, Iran.

2- Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China.

3- Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran.

mtourandaz@rouzbahan.ac.ir

Abstract

Predicting the change of convert Mild Cognitive Impairment (MCI) to Alzheimer's disease (AD) is an important issue to prevent Alzheimer's disease. This study proposes a new classification method, which is used sparse coding for selecting the useful features of biological markers in the human brain, for improving the accuracy of prediction of the infecting with Alzheimer's disease in MCI subjects. In the proposed method, after features which can improve the accuracy of classification are selected, predicting the infecting with AD will be done by the multi-kernel based learning methods. The empirical results, which applied on MCI subjects' data sets, show that the proposed method significantly improves the accuracy of the prediction of AD in comparison with previous methods.

Keywords: Classification, Sparse coding, Support Vector Machine, MRI, PET, CSF

1- M.Sc. Student, Artificial Intelligence.

2- Research Assistant, Artificial Intelligence.

3- Ph.D. Student, Electrical Engineering.

ارائه روشی جدید مبتنی بر انتخاب ویژگی های داده با استفاده از کد گذاری تُنک جهت پیش بینی بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی

ارائه دهنده مقاله: مرتضی تورانداز^۱

مرتضی تورانداز^۱، محمد یوسف نژاد^۲، صابر نوریان^۳

۱- دانشکده علوم کامپیوتر و هوش مصنوعی، موسسه آموزش عالی روزبهان ساری، ایران.

۲- دانشکده علوم و تکنولوژی کامپیوتر، دانشگاه هوا و فضای نانجینگ، چین.

۳- دانشکده برق و الکترونیک، دانشگاه علم و صنعت ایران، ایران.

ایمیل مسئول مقاله: (mtourandaz@rouzbahan.ac.ir)

چکیده

پیش بینی ابتلا به بیماری آلزایمر در افرادی که مبتلا به اختلالات خفیف شناختی هستند راه حلی مهم برای پیش گیری از ابتلای این افراد به این بیماری است. هدف از ارائه این مقاله، پیشنهاد روشی جدید جهت بهبود دقت طبقه بندی به منظور پیش بینی بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی بر اساس استفاده از کد گذاری تُنک در انتخاب ویژگی های مفید از نشانگرهای زیستی مختلف مغز انسان است. در این روش ابتدا با استفاده از کد گذاری تُنک ویژگی هایی که دقت طبقه بندی را کاهش می دهند حذف شده، سپس طبقه بندی با استفاده از روش های مبتنی بر یادگیری چند هسته ای، انجام می پذیرد. داده های مورد بحث در این مقاله شامل افرادی است که به اختلالات خفیف شناختی دچار هستند. نتایج نشان می دهد که دقت طبقه بندی با استفاده از روش پیشنهادی نسبت به روش های دیگر بهبود پیدا خواهد کرد.

واژه های کلیدی: طبقه بندی، کد گذاری تُنک، ماشین بردار پشتیبان، MRI، PET، CSF

۱- دانشجوی کارشناسی ارشد، هوش مصنوعی.

۲- دستیار پژوهش، هوش مصنوعی.

۳- دانشجوی دکتری، برق و الکترونیک.

مقدمه

در حال حاضر بیماری آلزایمر^۱ یکی از شایع ترین دلایل زوال عقل در افراد سالخورده است. تحقیقات به عمل آمده نشان می دهد که تعداد افراد مبتلا به این بیماری تا ۲۰ سال آینده، دو برابر تعداد فعلی و در سال ۲۰۵۰ از هر ۸۵ نفر یک نفر به این بیماری مبتلا خواهد بود (Brookmeyer, Johnson, Ziegler, Arrighi, 2007). به دلیل اینکه احتمال ابتلا به بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی بالاست، پیش بینی ابتلا به بیماری آلزایمر در این افراد بسیار بااهمیت خواهد بود (Thompson, Apostolova, 2007). در حال حاضر از نشانگرهای زیستی مختلفی مانند ۱- مقدار نرون های از دست رفته نواحی مختلف مغز و اتصالات تخریب شده ی بین آنها^۲ که از تصاویر ساختاری MRI^۳ مغز قابل استخراج است (De Leon et al, 2007) -۲ هیپومتابولیسیم اندازه گیری شده به وسیله تصویربرداری کارکردی^۴ (Morris et al, 2001) و ۳- تعیین مقدار پروتئین مغز به وسیله مایع مغزی نخاعی^۵ (Bouwman et al, 2009) و (Mattsson et al, 2009) برای تشخیص اختلالات به وجود آمده در مغز به دلیل ابتلا به بیماری های شناختی مانند آلزایمر و اختلالات خفیف شناختی^۶ استفاده می شود. تحقیقات نشان می دهد ویژگی های موثر برای تشخیص بیماری های شناختی با استفاده از نشانگرهای زیستی حاصل از تصاویر MRI باید از مناطقی از مغز که به دلیل ابتلا به این بیماری تخریب شده اند استخراج گردد (Chetelat et al, 2002) و (Convit et al, 2000). با استفاده از اطلاعات حاصل از تصویربرداری کارکردی PET به عنوان نشانگر زیستی، میزان کاهش متابولیسیم گلوکز مناطق مغز در افرادی که مبتلا به بیماری شناختی هستند تعیین می گردد. مایع مغزی نخاعی به عنوان نشانگر زیستی دیگر، میزان پروتئین موجود در نواحی مختلف مغز را نشان می دهد. در اغلب کارهای انجام شده تنها از اطلاعات حاصل از یک نشانگر زیستی برای طبقه بندی افراد مبتلا به بیماری آلزایمر و افراد سالم استفاده شده است (Cuingnet et al, 2009) و (Notestine et al, 2009). این در حالی است که تحقیقات اخیر نشان می دهد ترکیب ویژگی های حاصل از نشانگرهای زیستی مختلف، قدرت تشخیص بیماری های شناختی را به مراتب افزایش می دهد. در واقع نشانگرهای زیستی مختلف، اطلاعاتی تولید می کنند که مکمل یکدیگرند (de Leon et al, 2007) و (Apostolova et al, 2010). اخیراً به منظور ترکیب اطلاعات حاصل از نشانگرهای زیستی مختلف، از ابزارهای یادگیری ماشین استفاده شده است (Zhang, Wang, Zhou, Yuan, Shen, 2011)، (Ye, Wu, Li, Chen, 2007)، (Escudero et al,)، (Wee et al, 2012) و (2013). این ابزارها در تشخیص بیماری های شناختی با استفاده از اطلاعات حاصل از نشانگرهای زیستی مختلف بسیار موثر هستند. هدف از ارائه این مقاله، پیشنهاد روشی جدید مبتنی بر استفاده از کدگذاری تَنک^۷ به منظور حذف ویژگی های غیرمفید جهت بالا بردن دقت طبقه بندی با هدف پیش بینی بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی است. در این مقاله از داده هایی استفاده می شود که نشان دهنده ویژگی های استخراج شده از سه نشانگر زیستی MRI، PET و CSF، مربوط به افراد مبتلا به اختلالات خفیف شناختی است که گروهی از آنها بعد از ۱۸ ماه به بیماری آلزایمر مبتلا شده و مابقی بعد از طی این مدت به آلزایمر مبتلا نشده اند.

در ادامه مقاله ابتدا در بخش دوم به بررسی مبانی نظری این مقاله خواهیم پرداخت، در بخش سوم پیشینه ای از تحقیقات و کارهای انجام شده را مرور خواهیم کرد. در بخش چهارم به بررسی و تحلیل روش پیشنهادی خواهیم پرداخت و در بخش پنجم

1 Alzheimer's disease

2 Atrophy

3 Brain structural imaging

4 Brain functional imaging

5 Cerebrospinal fluid (CSF)

6 Mild cognitive impairment (MCI)

7 Sparse coding

داده‌های مورد استفاده در این مقاله و نتایج به دست آمده را شرح خواهیم داد. در بخش آخر نیز نتیجه‌گیری و خط و مشی کارهای آتی بیان خواهد شد.

مبانی نظری پژوهش

استفاده از اطلاعات حاصل از نشانگرهای زیستی مختلف یکی از چالش‌برانگیزترین مسائل در تشخیص بیماری‌های شناختی است (Apostolova et al, 2010)، (de Leon et al, 2007)، (Walhovd et al, 2010) و (Landao et al, 2010). در مطالعات اخیر، اطلاعات حاصل از سه نشانگر زیستی MRI، PET و CSF برای عمل طبقه‌بندی افراد مبتلا به آلزایمر و افراد سالم مورد استفاده قرار گرفته است (Zhang et al, 2011) و (Hinrichs, Singh, Xu, Johnson, 2011). برای ترکیب این اطلاعات، از روش‌های مبتنی بر یادگیری چند هسته‌ای^۱ استفاده می‌شود (Hinrichs, Singh, Xu, Johnson, 2009)، (Lanckriet, Deng, Cristianini, Jordan, 2004) و (Wang, Chen, Sun, 2008). روش‌های مبتنی بر یادگیری چند هسته‌ای از ایده ادغام اطلاعات حاصل از منابع اطلاعاتی مختلف برای بهبود بخشیدن به فهم داده‌ها و در نهایت بالا بردن قدرت طبقه‌بندی بر اساس آن‌ها، بهره می‌برند. با استفاده از ابزار یادگیری ماشین مانند ماشین بردار پشتیبان^۲ می‌توان با استفاده از اطلاعات به دست آمده از روش‌های مبتنی بر یادگیری چند هسته‌ای، عمل طبقه‌بندی را انجام داد. انتخاب ویژگی‌ها یکی از پرکاربردترین و چالش‌برانگیزترین زمینه‌ها برای بهبود نتایج طبقه‌بندی داده‌های مختلف به حساب می‌آید. انتخاب ویژگی بر اساس کدگذاری^۳، ویژگی‌هایی را که برای انجام عمل طبقه‌بندی مفید نیستند، حذف می‌کند، که این کار موجب بالا رفتن دقت عمل طبقه‌بندی می‌گردد.

پیشینه پژوهش

برای تشخیص بیماری‌های شناختی مطالعات بسیاری انجام و روش‌های زیادی ارائه شده است. در اکثر تحقیقات، اطلاعات حاصل از تصویربرداری ساختاری و کارکردی مغز برای تشخیص این بیماری‌ها مورد مطالعه قرار گرفته است (Wee et al, 2012). در تحقیقات و کارهای اولیه، تشخیص این بیماری بر اساس استفاده از ویژگی‌های حاصل از تنها یک نشانگر زیستی بوده است. اخیراً بر اساس مطالعات انجام‌گرفته، دانشمندان دریافته‌اند که اطلاعات حاصل از ترکیب ویژگی‌های استخراج شده از نشانگرهای زیستی مختلف موجب بالا رفتن دقت در تشخیص بیماری آلزایمر می‌شود. وموری (Vemuri) و همکاران، تنها از نشانگرهای زیستی MRI و CSF برای تشخیص بیماری‌های شناختی بهره برده‌اند (Vemuri et al, 2007). فلگیبل (Fellgiebel) و همکاران از دو نشانگر زیستی PET و CSF برای تشخیص این بیماری‌ها (Fellgiebel, Scheurich, Bartenstein, Muller, 2007) و ولهوود (Walhovd) و همکاران در تحقیقاتشان از هر سه نشانگر زیستی PET، MRI و CSF برای تشخیص بیماری‌های شناختی استفاده کرده‌اند (Walhovd et al, 2010). جانگ (Zhang) و همکاران با استفاده از ویژگی‌های استخراج شده از سه نشانگر زیستی PET، MRI و CSF روشی مبتنی بر یادگیری چند هسته‌ای ارائه داده‌اند (Zhang et al, 2011). جانگ و همکاران نشان داده‌اند که نتایج به دست آمده بر اساس ترکیب ویژگی‌های استخراج شده از نشانگرهای زیستی مختلف به مراتب بهتر از نتایج به دست آمده با استفاده از تنها یک نشانگر زیستی است. داواتسیکاس (Davatzikos) و همکاران از نشانگرهای زیستی MRI و CSF برای پیش‌بینی تبدیل بیماری اختلالات خفیف شناختی به بیماری آلزایمر بهره برده‌اند (Davatzikos, 2011).

1 Multikernel based approach

2 Support Vector Machine(SVM)

بایانو و همکاران (Biao) با استفاده از تصاویر کارکردی مغز و تشکیل شبکه ارتباطی بین نواحی مشخص شده از مغز توانسته اند روشی جدید برای تشخیص بیماری های شناختی ارائه دهند. آن ها با استفاده از روش های تحلیل شبکه های ارتباطی، ویژگی های محلی^۱ و توپولوژیک^۲ شبکه را استخراج و بعد از ساخت ماتریس هسته^۳ هر کدام، آن ها را با کمک یادگیری چند هسته ای با هم ادغام کرده و به روشی جدید و با دقت بالا برای تشخیص بیماری های شناختی رسیده اند (Biao et al, 2014).

وی (Wee) و همکاران با به کارگیری کدگذاری تَنک روی شبکه های کارکردی حاصل از تصاویر کارکردی مغز روشی جدید برای تشخیص بیماری آلزایمر ارائه داده اند (Wee, Yap, Zhang, Wang, Shen, 2012). آن ها در این روش برای ساخت شبکه کارکردی مغز از کدگذاری تَنک به منظور حذف ارتباطات غیرمفید استفاده کرده اند. وی و همکاران نشان داده اند که استفاده از کدگذاری تَنک در ساخت شبکه کارکردی مغز منجر به بالا رفتن دقت در تشخیص بیماری های شناختی می شود.

روش پژوهش

هدف از ارائه این مقاله بالا بردن دقت طبقه بندی افراد مبتلا به اختلالات خفیف شناختی بر اساس ویژگی های استخراج شده از نشانگرهای زیستی PET، MRI و CSF با هدف پیش بینی ابتلای این افراد به بیماری آلزایمر است. نکته کلیدی در این مقاله استفاده از کدگذاری تَنک برای انتخاب ویژگی های مفید از میان ویژگی هایی است که از نشانگرهای زیستی فوق استخراج می گردند. با استفاده از ویژگی های انتخاب شده و بر اساس یادگیری چند هسته ای، ماشین بردار پشتیبان جهت طبقه بندی گروه های مختلف مبتلا به اختلالات خفیف شناختی، آموزش داده می شود. نتایج نشان می دهد که استفاده از کدگذاری تَنک برای حذف ویژگی های غیرمفید موجب بالا رفتن دقت این طبقه بندی می گردد.

طبقه بندی چند هسته ای

در این بخش یک مرور کلی روی روش های یادگیری مبتنی بر هسته خواهیم داشت. این روش ها به راحتی قابل اعمال به ماشین بردار پشتیبان جهت طبقه بندی داده ها هستند (Scholkopf, Smola, 2002). قبل از پرداختن به روش های یادگیری چند هسته ای، ابتدا ماشین بردار پشتیبان تک هسته ای را مورد بررسی قرار می دهیم. ایده اصلی ماشین بردار پشتیبان مبتنی بر هسته به صورت خلاصه بدین شرح است: ابتدا داده هایی که به صورت خطی جدایی پذیر نیستند از فضای اصلی به فضای بالاتر^۴ که در آن داده ها به صورت خطی جدایی پذیر باشند نگاشت می شوند. این نگاشت از طریق توابع هسته^۵ قابل اعمال است. سپس مدل مورد نظر با بیشینه مقدار حاشیه برای طبقه بندی داده ها ساخته می شود.

برای ساخت مدل بر اساس ترکیب اطلاعات حاصل از ویژگی های استخراج شده از نشانگرهای زیستی MRI، PET و CSF جهت تشخیص یا پیش بینی بیماری های شناختی از ماشین بردار پشتیبان چند هسته ای^۶ استفاده می شود (Zhang et al, 2011). برای بررسی ماشین بردار پشتیبان چند هسته ای فرض کنید n داده آموزشی و m وجه (به طور مثال نشانگر زیستی) داریم. $X_i^{(m)}$

-
- 1 Local feature
 - 2 Topological feature
 - 3 Kernel matrix
 - 4 Feature space
 - 5 Kernel function
 - 6 Multikernel Support Vector Machine (MKLSVM)

نشان دهنده بردار ویژگی m -امین وجه از i -امین داده آموزشی است و برچسب معادل آن y_i دارای یکی از دو مقدار -1 و 1 است. ماشین بردار پشتیبان چند هسته ای تابع هدف در رابطه (۱) را بهینه می کند.

$$\min_{w^{(m)}, b, \xi_i} \frac{1}{2} \sum_{m=1}^M \beta_m \|w^{(m)}\|^2 + c \sum_{ij} \xi_i \quad (1)$$

$$s. t \ y_i \left(\sum_{m=1}^M \beta_m \left((w^{(m)})^T \phi^{(m)}(x_i^{(m)}) + b \right) \right) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, \dots, n$$

در رابطه (۱)، $w^{(m)}$ ، $\phi^{(m)}$ و $\beta_m \geq 0$ به ترتیب نشان دهنده بردار وزن، تابع هسته نگاشت و وزن هر وجه برای ترکیب اطلاعات حاصل از ویژگی های استخراج شده است. فرم دوگان ماشین بردار پشتیبان همانند فرم اولیه آن به دنبال بهینه کردن تابع معیاری است که در رابطه (۲) آورده شده است.

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j \sum_{m=1}^M \beta_m k^{(m)}(x_i^{(m)}, x_j^{(m)}) \quad (2)$$

$$s. t \ \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, n$$

در رابطه (۲)، $k^{(m)}(x_i^{(m)}, x_j^{(m)}) = \phi^{(m)}(x_i^{(m)})^T \phi^{(m)}(x_j^{(m)})$ ، برای دو داده آموزشی x_i و x_j روی m -امین وجه از داده است. برای داده آزمایشی جدید $x = \{x^{(1)}, x^{(2)}, \dots, x^{(M)}\}$ مقدار $k^{(m)}(x_i^{(m)}, x^{(m)}) = \phi^{(m)}(x_i^{(m)})^T \phi^{(m)}(x^{(m)})$ که نشان دهنده مقدار تابع هسته مورد استفاده میان داده آزمایشی و تمام داده های آموزشی روی m -امین وجه است، محاسبه می گردد. سپس رابطه (۳) برای پیش بینی برچسب کلاسی که داده آزمایشی به آن تعلق دارد اعمال می گردد.

$$f(x^{(1)}, x^{(2)}, \dots, x^{(M)}) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i \sum_{m=1}^M \beta_m k^{(m)}(x_i^{(m)}, x_j^{(m)} + b) \right) \quad (3)$$

اگر $k(x_i, x_j) = \sum_{m=1}^M \beta_m k^{(m)}(x_i^{(m)}, x_j^{(m)})$ را به عنوان تابع هسته ادغام شده بین دو داده ای آموزشی x_i و x_j و $k(x_i, x) = \sum_{m=1}^M \beta_m k^{(m)}(x_i^{(m)}, x^{(m)})$ را به عنوان تابع هسته ادغام شده بین داده ای آموزشی x_i و داده ای آزمایشی x در نظر بگیریم، آنگاه می توان ماشین بردار پشتیبان چند هسته ای را به وسیله ی ماشین بردار تک هسته ای پیاده سازی کرد. در رابطه های فوق این محدودیت وجود دارد که $\sum_m \beta_m = 1$. می توان با قرار دادن مقادیر مختلف برای هر β_m در نظر گرفتن محدودیت ذکر شده و با استفاده از اعتبارسنجی k -تکه ای^۱ روی داده های آموزشی بهترین مقادیر را برای این متغیرها یافت. در این مقاله برای آموزش ماشین بردار پشتیبان از ابزار LIBSVM (Chang, Lin, 2001) استفاده شده است.

کدگذاری تُنک

نکته کلیدی در این مقاله استفاده از کدگذاری تُنک جهت انتخاب ویژگی های مفید برای پیش بینی بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی است. انتخاب ویژگی های مفید با بهینه کردن تابع معیار در رابطه (۴) امکان پذیر است.

$$\min_{w,b} \frac{1}{2} \sum_i^N (y_i - w^{(T)}x_i - b)^2 + \lambda \|w\|_1 \quad (4)$$

در رابطه (۴)، x_i نشان دهنده بردار ویژگی، y_i نشان دهنده برچسب کلاسی است که داده x_i به آن تعلق دارد، b مقدار عرض از مبدا داده x_i ، w ضرایب رگرسیون بردار ویژگی داده x_i ، λ پارامتر تنظیم و N تعداد داده آموزشی است. $\|w\|_1$ ، تنظیم کننده L_1 -norm است که برای کدگذاری تُنک در فضای ویژگی مورد استفاده قرار می گیرد. با استفاده از این روش ویژگی های معادل با ضرایبی که با بهینه کردن رابطه (۴) صفر می شوند، از بردار ویژگی حذف می گردند. بهترین مقدار برای پارامتر تنظیم ($\lambda > 0$) در رابطه (۴) با استفاده از عمل اعتبارسنجی k -تکه ای روی داده های آموزشی قابل محاسبه است. در این مقاله از ابزار LASSO (Tibshirani, 1996) برای بهینه کردن رابطه (۴) استفاده می شود.

در این مقاله ویژگی های غیرمفید استخراج شده از سه نشانگر زیستی MRI، PET و CSF برای تشخیص بیماری آلزایمر با استفاده از این روش حذف ویژگی، حذف شده و سپس با استفاده از ویژگی های باقیمانده هسته های مربوط به هر کدام از این ویژگی ها، محاسبه و باهم ادغام می شوند. در نهایت با استفاده از ماشین بردار پشتیبان و با استفاده از هسته های ترکیب شده، عمل طبقه بندی روی گروه های مختلف مبتلا به اختلالات خفیف شناختی به منظور پیش بینی ابتلا به بیماری آلزایمر در این افراد انجام می پذیرد. در شکل (۱) شمای کلی این مراحل به تصویر کشیده شده است.

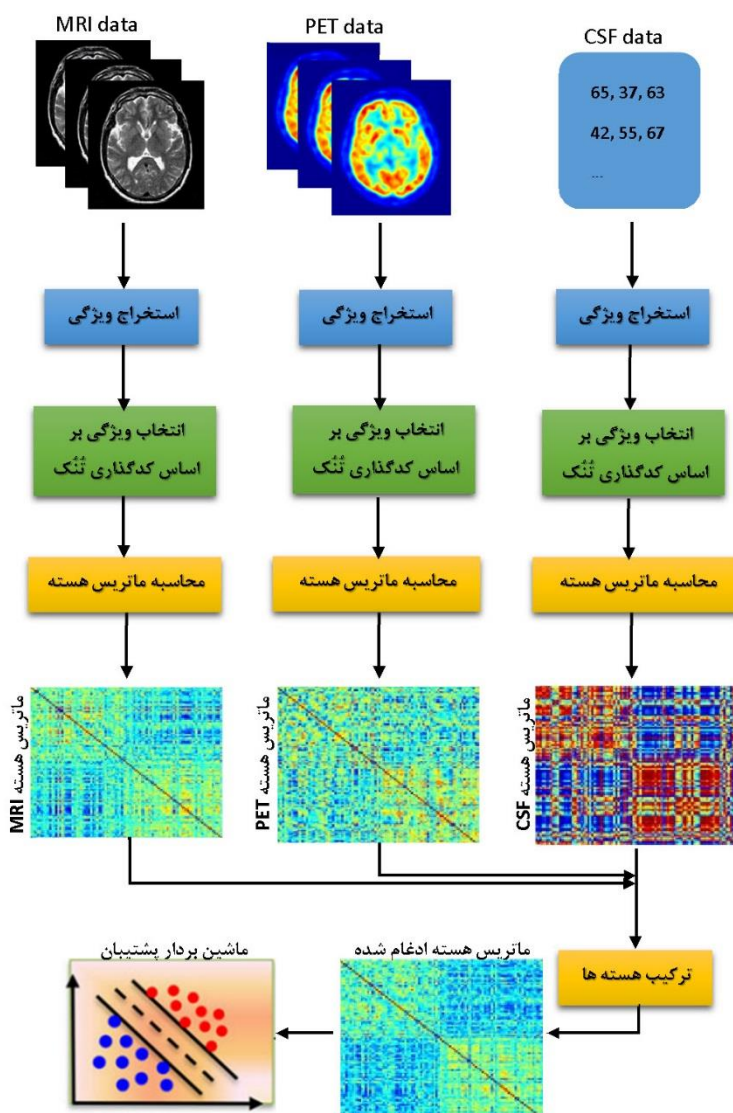
اعتبارسنجی

برای سنجیدن کارایی ماشین بردار پشتیبان آموزش داده شده با استفاده از روش پیشنهادی از اعتبارسنجی k -تکه ای بهره می بریم. برای این کار مجموعه داده های موجود را به ۱۰ زیرمجموعه افراز می کنیم. در هر بار از ۹ زیرمجموعه برای آموزش ماشین بردار پشتیبان و از یک زیرمجموعه باقیمانده برای سنجش مدل آموزش داده شده استفاده می کنیم. این کار به تعداد ۱۰ بار به منظور ساختن زیرمجموعه هایی با ترکیب داده های مختلف و اطمینان از رخ ندادن Overfitting مدل، تکرار می شود. برای یافتن وزن بهینه ماتریس هسته هر کدام از m نشانگر زیستی (β_m) به منظور ترکیب هسته ها، از داده های آموزشی استفاده می کنیم. بدین ترتیب که با قرار دادن مقادیر مختلف از ۰ تا ۱ و اندازه گام ۰٫۱ برای β_m ها و آموزش مدل بر اساس این مقادیر (با در نظر گرفتن محدودیت $\sum \beta_m = 1$)، دقت مدل را بر حسب داده های آزمایشی می سنجیم و وزن هایی را که موجب بیشترین دقت برای مدل شده اند، به عنوان وزن هسته ها در نظر می گیریم.

به منظور یکسان سازی محدوده مقادیر در ویژگی های مختلف استخراج شده از نشانگرهای زیستی ابتدا باید داده ها را نرمال نمود. برای این کار به صورت رابطه (۵) عمل می کنیم.

$$f_i = (f_i - \bar{f}_i) / \sigma_i \quad (5)$$

در رابطه (۵)، f_i ویژگی مورد نظر، \bar{f}_i مقدار میانگین مقادیر مختلف ویژگی f_i و σ_i مقدار انحراف معیار مقادیر ویژگی f_i است.



شکل ۱- نمایش شماتیک روند کلی کار برای طبقه بندی داده های چندوجهی^۱

خلاصه روش پیشنهادی

در شکل (۱) روند کلی کار در روش پیشنهادی به تصویر کشیده شده است، همان طور که در این شکل می بینید ابتدا ویژگی های مورد نیاز از سه نشانگر زیستی MRI، PET و CSF استخراج می گردند. تابع معیار نشان داده شده در رابطه (۴) با استفاده از ابزار LASSO به منظور انجام عمل کدگذاری تُنک جهت حذف ویژگی های غیرمفید بهینه می گردد. در مرحله بعد ماتریس هسته مربوط به هر کدام از نشانگرهای زیستی فوق با استفاده از ویژگی های باقیمانده محاسبه شده و با یافتن بهترین اوزان ماتریس های هسته به دست آمده، این ماتریس ها باهم ترکیب و ماشین بردار پشتیبان جهت انجام عمل طبقه بندی آموزش داده می شود. در شکل (۲) شبه کد مربوط به روش پیشنهادی آورده شده است. ورودی تابع تعریف شده، ویژگی های استخراج شده از نشانگرهای زیستی MRI، PET

و CSF و خروجی آن ماشین بردار پشتیبان آموزش داده شده جهت پیش بینی بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی است.

Pseudocode for the proposed method

Inputs: Feature vector of MRI biomarker
Feature vector of PET biomarker
Feature vector of CSF biomarker
Output: SVM model for predict Alzheimer's disease
Function LMCI_HMCI_SVM (Inputs)

1. Select useful features with sparse coding
2. Compute kernel matrixes
3. Compute weight of each kernel matrix
4. Combine all matrixes
5. Train SVM model for predict Alzheimer's disease in MCI subjects via K-fold cross validation
6. Output=SVM model

End function

شکل ۲- شبه کد مربوط به روش پیشنهادی

ارزیابی

در این بخش به بررسی داده های مورد استفاده در این مقاله و نتایج به دست آمده در پیاده سازی روش پیشنهادی می پردازیم. برای ارزیابی نتایج به دست آمده با روش جدید، آن ها را با نتایج روش های رایج برای طبقه بندی مانند شبکه عصبی، درخت تصمیم و ... مقایسه می کنیم.

تجزیه و تحلیل داده ها

داده های مورد استفاده در این مقاله از پایگاه های داده آزمایشگاه تصویربرداری عصبی بیماری آلزایمر^۱ (ADNI) استخراج شده است.^۲ در این آزمایشگاه از مغز افراد مختلف تصویربرداری می شود. افرادی که در این آزمایشگاه مورد تصویربرداری عصبی قرار می گیرند دارای سنی بین ۵۵ و ۹۰ سال بوده و عبارت اند از ۱- افراد سالم^۳، ۲- افراد مبتلا به اختلالات خفیف شناختی و ۳- افراد مبتلا به بیماری آلزایمر. در این مقاله از اطلاعات مربوط به ۹۹ فرد مورد آزمایش در ADNI که هر سه نشانگر زیستی PET، MRI و CSF برای آن ها محاسبه شده، استفاده می کنیم که عبارت اند از: ۴۳ فرد مبتلا به اختلالات خفیف شناختی که بعد از ۱۸ ماه به بیماری آلزایمر مبتلا شده اند و ۵۶ فرد دیگر نیز مبتلا به اختلالات خفیف شناختی هستند با این تفاوت که بعد از ۱۸ ماه به بیماری آلزایمر مبتلا نشده اند.

برای استخراج ویژگی های مربوط به نشانگرهای زیستی حاصل از MRI و PET تصاویر مغز حاصل از تصویربرداری ساختاری و کارکردی به ۹۳ ناحیه^۴ تقسیم می شود. برای تصاویر ساختاری، سطح ماده خاکستری هر ناحیه از مغز به عنوان ویژگی آن در نظر گرفته شده است. برای تصاویر کارکردی بعد از اعمال پیش پردازش هایی همچون حذف حرکات اضافی سر^۵، اصلاح زمانی سری زمانی تصاویر^۶ برای هر فرد و اصلاح مکانی^۷ هر کدام از تصاویر موجود در سری زمانی نسبت به اولین تصویر موجود در این سری زمانی

1 Alzheimer's Disease Neuroimaging Initiative

2 www.loni.ucla.edu/ADNI

3 Healthy subjects

4 Region of interest

5 Motion correction

6 Slice timing

7 Realignment

برای هر فرد، مقدار میانگین شدت هر ناحیه به عنوان ویژگی آن در نظر گرفته می شود. برای نشانگر زیستی CSF نیز سه مقدار CSF p-tau، CSF t-tau، Aβ42 استفاده شده است (Zhang et al, 2011)؛ بنابراین برای هر فرد در مجموع ۱۸۹ ویژگی شامل ۹۳ ویژگی از تصویر MRI، ۹۳ ویژگی از تصویر PET و ۳ ویژگی از مایع مغزی نخاعی در نظر گرفته می شود.

نتایج

برای انجام مقایسه و سنجیدن دقت روش پیشنهادی، روش های درخت تصمیم^۱، شبکه عصبی^۲، K-نزدیک ترین همسایه^۳ و ماشین بردار پشتیبان را نیز به منظور طبقه بندی داده های مورد نظر اجرا کرده ایم که نتایج در جدول (۱) نشان داده شده است. روش پیشنهادی در محیط MATLAB پیاده سازی و نتایج آزمایش روی میانگین و خطای استاندارد ۱۰ بار آموزش و آزمایش مدل در جدول (۱) گزارش شده است. آموزش مدل با استفاده از اعتبارسنجی ۱۰-تکه ای انجام شده است، به این صورت که داده های موجود به ۱۰ قسمت تقسیم شده و در هر مرحله ۹ قسمت از داده ها برای آموزش مدل و ۱ قسمت باقیمانده برای سنجیدن آن به کار می رود. این فرآیند برای هر روش آزمایش (درخت تصمیم، شبکه عصبی و ...) به تعداد ۱۰ بار اجرا شده و دقت مدل بر اساس میانگین و با در نظر گرفتن خطای استاندارد^۴ اندازه گیری شده است. سطرهای ۱ تا ۳ این جدول نتایج به دست آمده از اجرای روش های گفته شده با استفاده از هر یک از نشانگرهای زیستی، در سطرهای چهارم تا ششم، نتایج با استفاده از ترکیب های مختلف دوتایی از ۳ نشانگر زیستی و در سطر آخر نیز نتایج بر اساس ترکیب هر سه این نشانگرها نشان داده شده است. برای محاسبه دقت آورده شده در جدول (۱) برای نشانگرهای زیستی MRI و PET، هر کدام ۹۳ ویژگی از ۹۳ ناحیه مغز و برای نشانگر زیستی CSF، ۳ ویژگی از مایع مغزی نخاعی در نظر گرفته شده است.

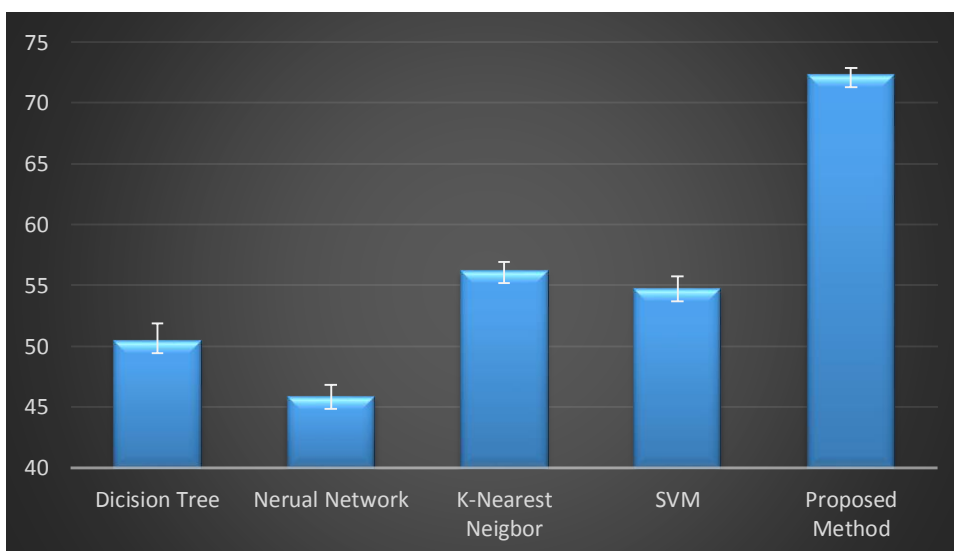
جدول ۱- نتایج حاصل از اجرای روش های طبقه بندی روی داده ها بر حسب میانگین دقت ± خطای استاندارد در ۱۰ بار اجرای مستقل

روش پیشنهادی	ماشین بردار پشتیبان	K-نزدیک ترین همسایه	شبکه عصبی	درخت تصمیم	نشان گر زیستی
۷۴/۸۴±۰/۸۸	۵۲/۸۹±۱/۰۷	۵۲/۸۴±۰/۷۲	۴۵/۱۶±۱/۰۱	۵۱/۹۴±۱/۶۴	MRI
۷۱/۴±۰/۷۵	۵۰/۷±۰/۹۴	۵۰/۱۳±۰/۶۲	۴۶/۰۱±۱/۳	۵۲/۴۵±۱/۳	PET
۵۸/۶۱±۰/۱۶	۵۶/۵۸±۰/۹۶	۵۶/۳۷±۰/۷	۴۵/۸۳±۰/۴۷	۵۱/۸±۱/۳۳	CSF
۷۷/۱۲±۰/۷۹	۵۷/۸۳±۰/۹۹	۵۸/۲۹±۰/۹۹	۴۵/۳۶±۰/۸۱	۴۶/۷۷±۱/۵	MRI+PET
۷۳/۶۴±۰/۶۳	۵۶/۳۷±۱/۰۴	۵۷/۷۱±۰/۷۵	۴۶/۶۴±۱/۳۹	۵۱/۶±۱/۱	MRI+CSF
۷۲/۸۶±۰/۵۸	۵۵/۳۲±۱/۲۸	۶۰/۴۲±۰/۵۹	۴۶/۳۳±۱/۲۹	۵۱/۵۳±۱/۵۹	CSF+PET
۷۷/۲۹±۰/۶۹	۵۳/۲۴±۱/۱۵	۵۷/۵۳±۱	۴۵/۶۷±۰/۶۵	۴۷/۱۳±۱/۴۳	MRI+PET+CSF

- 1 Decision Tree (DT)
- 2 Neural Network (NN)
- 3 K-nearest neighbor (KNN)
- 4 Standard error

نتایج حاصل از روش پیشنهادی در آخرین ستون آورده شده است. برای محاسبه دقت مدل بر اساس روش پیشنهادی و در حالتی که تنها از ویژگی های یک نشانگر زیستی استفاده می شود، ابتدا با استفاده از کد گذاری تُنک ویژگی های غیرمفید حذف شده و در نهایت ماشین بردار پشتیبان را با ویژگی های باقی مانده آموزش می دهیم. همچنین در حالتی که از ترکیب های مختلف این نشانگرهای زیستی استفاده می شود، ابتدا ویژگی های غیرمفید با استفاده از کدگذاری تُنک، حذف و در مرحله بعد ماتریس هسته هر کدام از نشانگرهای زیستی محاسبه شده و در نهایت ماشین بردار پشتیبان با استفاده از ترکیب این ماتریس ها آموزش داده می شود. نتایج نشان می دهند که روش پیشنهادی موجب بهبود دقت طبقه بندی افراد مبتلا به اختلالات خفیف شناختی می گردد.

در شکل (۳) نمودار مربوط به میانگین دقت و خطای استاندارد به دست آمده در پیاده سازی روش های گفته شده آورده شده است. این نمودار نشان می دهد که روش پیشنهادی دارای بیشترین دقت و کمترین خطای استاندارد است.



شکل ۳- نمودار میانگین درصد صحت و خطای استاندارد روش های مختلف

نتیجه گیری

در این مقاله طرحی جدید مبتنی بر استفاده از کدگذاری تُنک به منظور انتخاب ویژگی های مفید از میان ویژگی های استخراج شده از نشانگرهای زیستی مغز انسان، برای بالا بردن دقت پیش بینی ابتلا به بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی ارائه شده است. این روش موجب بالا رفتن دقت پیش بینی بیماری آلزایمر نسبت به روش های قبلی می شود. در روش اولیه ویژگی های استخراج شده از نشانگرهای زیستی مختلف به صورت یکجا در نظر گرفته شده و عمل طبقه بندی روی داده ها به کمک این ویژگی ها انجام می گیرد. اخیراً با استفاده از روش های یادگیری چند هسته ای، برای هر کدام از نشانگرهای زیستی موجود، هسته های جداگانه در نظر گرفته شده و با استفاده از ادغام این هسته ها عمل طبقه بندی صورت می پذیرد. نتایج به دست آمده با استفاده از داده های مربوط به افراد مبتلا به اختلالات خفیف شناختی نشان می دهد که روش پیشنهادی این مقاله مبتنی بر انتخاب ویژگی های مفید برای انجام عمل طبقه بندی به صورت چشمگیری موجب بالا رفتن دقت پیش بینی ابتلا به بیماری آلزایمر در افراد مبتلا به اختلالات خفیف شناختی می گردد.

تاکنون مطالعات بسیار زیادی برای تشخیص بیماری های شناختی انجام شده و روش های بسیار متنوعی برای بالا بردن قدرت تشخیص این بیماری ها ارائه شده است. از جمله این روش ها ساخت شبکه کارکردی از روی تصاویر کارکردی مغز جهت تشخیص بیماری های شناختی است. در کارهای آتی به بررسی روش های مختلف استخراج ویژگی با استفاده از مفاهیم موجود در تحلیل شبکه های ارتباطی جهت استخراج ویژگی های موثر از شبکه های کارکردی مغز در تشخیص بیماری های شناختی خواهیم پرداخت.

منابع

- Apostolova, L. G., Hwang, K. S., Andrawis, J. P., Green, A. E., Babakchianian, S., Morra, J. H., Cummings, J. L., Toga, A. W., Trojanowski, J. Q., Shaw, L. M., Jack, C. R. Jr., Petersen, R. C., Aisen, P. S., Jagust, W. J., Koeppe, R. A., Mathis, C. A., Weiner, M. W., Thompson, P. M. (2010). "3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects". *Neurobiol Aging*. 31:1284–1303.
- Biao, J., Zhang, D., Gao, W., Wang, Q., Wee, C. Y., Shen, D. (2014). "Integration of Network Topological and Connectivity Properties for Neuroimaging Classification" *IEEE transactions on biomedical engineering*, vol. 61, no. 2.
- Brookmeyer, R, Johnson, E, Ziegler-Graham, K, and Arrighi, H. M. (2007). "Forecasting the global burden of Alzheimer's disease," *Alzheimer's Dementia*, vol. 3, pp. 186–191.
- Bouwman, F. H., van der Flier, W. M., Schoonenboom, N. S., van Elk, E. J., Kok, A., Rijmen, F., Blankenstein, M. A., Scheltens, P. (2007). "Longitudinal changes of CSF biomarkers in memory clinic patients". *Neurology*.69:1006–1011.
- Chang, C. C., Lin, C. J. (2001). "LIBSVM: a library for support vector machines".
- Chetelat, G., Desgranges, B., de la Sayette, V., Viader, F., Eustache, F., Baron, J. C. (2002). "Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment". *Neuroreport*. 13:1939–1943.
- Convit, A., de Asis, J., de Leon, M. J., Tarshish, C. Y., De Santi, S., Rusinek, H. (2000). "Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease". *Neurobiology of Aging*. 21:19–26.
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M. O., Chupin, M., Benali, H., Colliot, O. (2010). "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database". *Elsevier, Neuroimage*. Volume 56, Issue 2, Pages 766–781.
- Davatzikos, C., Priyanka, B., Leslie, M., Kayhan, N., Batmangheich, N., Trojanowski, Q. J. (2010). "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, pattern classification". *Neurobiol aging*. 2011 Dec;32(12):2322.
- de Leon, M. J., Mosconi, L., Li, J., De Santi, S., Yao Y, Tsui, W. H., Pirraglia, E., Rich, K., Javier, E., Brys, M., Glodzik, L., Switalski, R., Saint Louis, L., A., Pratico, D. (2007). "Longitudinal CSF isoprostane and MRI atrophy in the progression to AD". *J Neurol*. 254:1666–1675.
- Escudero, J., Ifeachor, E., Zajicek, J. P., Green, C., Shearer, J., Pearson, S. (2013). "Machine learning-based method for personalized and cost-effective de-tetection of Alzheimer's disease". *IEEE Trans. Biomed. Eng.*, vol. 60, no.1, pp. 164–168
- Fellgiebel, A., Scheurich, A., Bartenstein, P., Muller, M. J. (2007). "FDG-PET and CSF phospho-tau for prediction of cognitive decline in mild cognitive impairment". *Psychiatry Res*. 155:167–171.
- Hinrichs, C., Singh, V., Xu, G. F., Johnson, S. C. (2009). "MKL for robust multi-modality AD classification". *Med Image Comput Assist Interv*. 12:786–794.
- Hinrichs, C., Singh, V., Xu, G. F., Johnson S. C. (2011). "Predictivemarkers forAD in multi-modality framework: An analysis of MCI progression in the ADNI population," *Neuroimage*, vol. 55, pp. 574–589, Mar. 15.
- Landkriet G. R., Deng, M., Cristianini, N., Jordan, M. I. (2004). "Noble WS. Kernel-based data fusion and its application to protein function prediction in yeast". *Pac Symp Biocomput*. 300–311.
- Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M., Foster, N. L., Aisen, P. S., Petersen, R. C., Shaw, L. M., Trojanowski, J. Q., Jack, C. R. Jr., Weiner, M. W., Jagust, W. J. (2010). "Comparing predictors of conversion and decline in mild cognitive impairment". *Neurology*. 2010;75:230–238.
- Mattsson, N., Zetterberg, H., Hansson, O., Andreasen, N., Parnetti, L., Jonsson, M., Herukka, S. K., van der Flier, W. M., Blankenstein, M. A., Ewers, M., Rich, K., Kaiser, E., Verbeek, M., Tsolaki, M., Mulugeta, E., Rosen, E., Aarsland, D., Visser, P. J., Schroder, J., Marcusson, J., de Leon, M., Hampel, H., Scheltens, P., Pirttila, T., Wallin, A., Jonhagen,

M. E., Minthon, L., Winblad, B., Blennow, K. (2009). "CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment". *Jama*.302:385–393.

Morris, J. C., Storandt, M., Miller, J. P., McKeel D. W., Price, J. L., Rubin, E. H., Berg, L. (2001). "Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease" *Archives of Neurology*.58:397–405.

Notestine, C., Roddey, J. C., Hagler, D. J. Jr., Holland. D., Karow, D. S, Pung, C. J., Brewer, J. B., Dale, A. M. (2009) "Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment". *Radiology*. 2009;251:195–205.

Scholkopf, B., Smola, A. (2002) "Learning with Kernels". The MIT Press.

Thompson, P. M., Apostolova L. G. (2007), "Computational anatomical methods as applied to ageing and dementia". *Brit. J. Radiol.*, vol. 80, pp. S78–S91.

Tibshirani, R. (1996). "Regression shrinkage and selection via the Lasso". *J R Stat Soc Ser B* 58:267–288.

Vemuri, P., Wiste, H. J., Weigand, S. D., Shaw, L. M., Trojanowski, J. Q., Weiner, M. W., Knopman, D. S., Petersen, R. C., Jack, C. R. (2007). "MRI and CSF biomarkers in normal, MCI, and AD subjects". *AAN Enterprises*.

Walhovd, K. B., Fjell, A. M., Brewer, J., McEvoy, L. K., Fennema-Notestine, C., Hagler D. J. Jr., Jennings, R. G., Karow, D., Dale, A.M. (2010). "Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease". *AJNR Am J Neuroradiol*31:347–354.

Wang, Z., Chen, S., Sun, T. (2008). "MultiK-MHKS: a novel multiple kernel learning algorithm". *IEEE Trans Pattern Anal Mach Intell*. 30:348–353.

Wee, C. Y., Yap, P. T., Zhang D. Q., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., Wang, L. H., Shen, D. G. (2012) "Identification of MCI individuals using structural and functional connectivity networks". *Neuroimage*, vol. 59, pp. 2045–2056, Feb. 1.

Wee, C. Y., Yap, P. T., Zhang, D. Q., Wang, L., Shen, D. (2012). "Constrained Sparse Functional Connectivity Networks for MCI Classification". *MICCAI 2012, Part II, LNCS 7511*, pp. 212–219.

Ye, J. P., Wu, T., Li, J. Chen, K. W. (2007). "Machine learning approaches for the neuroimaging study of Alzheimer's disease". *Computer*, vol. 44, pp. 99–101, Apr. 2011.vol. 3, pp. 186–191.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D. (2011). "Multimodal classification of Alzheimer's disease and mild cognitive impairment". *Neuroimage*, vol. 55, pp. 856–867, Apr. 1.