

## ارائه روشی جدید جهت تشخیص بیماری آلزایمر مبتنی بر انتخاب ویژگی‌های داده با استفاده از کدگذاری تُنک

ارائه‌دهنده: مرتضی تورانداز<sup>۱</sup>

m.tourandaz@gmail.com

مرتضی تورانداز<sup>۱</sup>، محمد یوسف نژاد<sup>۲</sup>، صابر نوریان<sup>۳</sup>

### چکیده

در سال‌های اخیر تشخیص بیماری آلزایمر به یکی از چالش‌برانگیزترین مسائل در حوزه پزشکی تبدیل شده است. هدف از ارائه این مقاله، پیشنهاد روشی جهت بهبود دقت تشخیص بیماری آلزایمر مبتنی بر استفاده از کدگذاری تُنک در انتخاب ویژگی‌های مفید از نشانگرهای زیستی مختلف مغز انسان است. در این روش، ابتدا با استفاده از کدگذاری تُنک ویژگی‌هایی که در طبقه‌بندی موجب کاهش دقت خواهند شد، حذف می‌شوند و سپس تشخیص بیماری آلزایمر با استفاده از روش‌های مبتنی بر یادگیری چند هسته‌ای انجام می‌گیرد. داده‌های مورد بحث در این مقاله شامل افراد مبتلا به بیماری آلزایمر و افراد سالم می‌باشند. نتایج آزمایش نشان می‌دهد که دقت تشخیص بیماری آلزایمر با استفاده از روش پیشنهادی نسبت به روش‌های پیشین بهبود چشمگیری پیدا کرده است.

واژه‌های کلیدی: طبقه‌بندی، کدگذاری تُنک، ماشین بردار پشتیبان، MRI، PET، CSF

۱. دانشجوی کارشناسی ارشد، دانشکده علوم کامپیوتر و هوش مصنوعی، موسسه آموزش عالی روزبهان ساری، ایران.

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

۳. دانشجوی دکتری، دانشکده برق و الکترونیک، دانشگاه علم و صنعت ایران، ایران.

## **To propose a new method for diagnose Alzheimer's disease based on the selected features via sparse coding**

**Paper presenter: Morteza Tourandaz<sup>1</sup>**

m.tourandaz@gmail.com

**Morteza Tourandaz<sup>1</sup>, Muhammad Yousefnezhad<sup>2</sup>, Saber Nourian<sup>3</sup>**

### **Abstract**

In recent years, the diagnosis of Alzheimer's disease (AD) has become one of the most challenging problems in medical fields. This paper proposes a new classification method which is used sparse coding for selecting the useful properties of biological markers in the human brain for improving the accuracy of diagnosis for AD. In the proposed method, features of collected data sets, which can improve the accuracy of classification, are selected by using sparse coding. Furthermore, the multi-kernel based learning methods are used for diagnosis of AD. The data set will be discussed in this paper contains normal and AD subjects. The empirical results show that the proposed method significantly improves the accuracy of the diagnosis of AD in comparison with previous methods.

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

---

1. M.Sc. Student, Department of Computer Science and Artificial Intelligence, Rouzbahan University, Sari, Iran.

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

3. Ph.D. Student, Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran.

## مقدمه

در حال حاضر بیماری آلزایمر<sup>۱</sup> یکی از شایع‌ترین دلایل زوال عقل در افراد سالخورده است. تحقیقات به عمل آمده نشان می‌دهد که تعداد افراد مبتلا به این بیماری تا ۲۰ سال آینده به دو برابر تعداد فعلی خواهد رسید و همچنین در سال ۲۰۵۰ از هر ۸۵ نفر یک نفر به این بیماری مبتلا خواهد بود (Brookmeyer, Johnson, Ziegler, Arrighi, 2007). بنابراین تشخیص این بیماری در اوایل شکل‌گیری آن بسیار بااهمیت خواهد بود (Thompson, Apostolova, 2007). در حال حاضر از نشانگرهای زیستی مختلفی مانند ۱- مقدار نرون‌های از دست رفته نواحی مختلف مغز و اتصالات تخریب‌شده<sup>۲</sup> بین آن‌ها که از تصاویر ساختاری MRI مغز قابل استخراج است (De Leon et al, 2007) ۲- هیپومتابولیسیم اندازه‌گیری شده به وسیله تصویربرداری کارکردی<sup>۴</sup> (Morris et al, 2001) و ۳- تعیین مقدار پروتئین مغز به وسیله مایع مغزی نخاعی<sup>۵</sup> (Bouwman et al, 2009) و (Mattsson N et al, 2009) برای تشخیص اختلالات به وجود آمده در مغز به دلیل ابتلا به بیماری‌های شناختی مانند آلزایمر و اختلالات خفیف شناختی<sup>۶</sup> استفاده می‌شود. تحقیقات نشان می‌دهد که ویژگی‌های موثر برای تشخیص این بیماری با استفاده از نشانگرهای زیستی حاصل از تصاویر MRI باید از مناطقی از مغز که به دلیل ابتلا به این بیماری تخریب‌شده‌اند 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 آلزایمر هستند تعیین می‌گردد. مایع مغزی نخاعی به عنوان نشانگر زیستی دیگر، میزان پروتئین موجود در نواحی مختلف مغز را نشان می‌دهد. در اغلب کارهای انجام‌شده تنها از اطلاعات حاصل از یک نشانگر زیستی برای طبقه‌بندی 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 می‌دهد ترکیب ویژگی‌های حاصل از نشانگرهای زیستی مختلف، قدرت تشخیص بیماری‌های شناختی را به مراتب افزایش می‌دهد. در واقع نشانگرهای زیستی متفاوت، اطلاعاتی تولید می‌کنند 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, 2013) و (Wee et al, 2012). این ابزارها در تشخیص بیماری‌های شناختی با استفاده از اطلاعات حاصل از نشانگرهای زیستی مختلف بسیار موثر هستند. هدف از این مقاله، ارائه روشی جدید مبنی بر استفاده از کدگذاری<sup>۷</sup> به منظور حذف ویژگی‌های غیرمفید جهت بالا بردن دقت تشخیص بیماری آلزایمر است. در این مقاله از داده‌هایی استفاده می‌شود که نشان‌دهنده ویژگی‌های استخراج‌شده از سه نشانگر زیستی MRI، PET و CSF، مربوط به افراد مبتلا به بیماری آلزایمر، افراد مبتلا به اختلالات خفیف شناختی و افراد سالم است.

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

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

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 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 از منابع اطلاعاتی مختلف برای بهبود بخشیدن به فهم داده‌ها و در نهایت بالا بردن قدرت تشخیص بین داده‌های متفاوت و دسته‌بندی آن‌ها، بهره می‌برند. با استفاده از ابزار یادگیری ماشین مانند ماشین بردار پشتیبان<sup>۱</sup> 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 تُنک، ویژگی‌هایی که برای انجام عمل طبقه‌بندی مفید نیستند را حذف می‌کند که این کار موجب بالا رفتن دقت عمل طبقه‌بندی می‌گردد.

## پیشینه پژوهش

برای تشخیص بیماری‌های شناختی مانند آلزایمر 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 یک نشانگر زیستی بوده است. اخیراً بر اساس مطالعات انجام گرفته، دانشمندان دریافته‌اند که استفاده از اطلاعات حاصل از ترکیب ویژگی‌های این نشانگرهای زیستی موجب بالا رفتن دقت در تشخیص بیماری آلزایمر می‌شود. وموری و همکاران (Vemuri et al, 2007) تنها از نشانگرهای زیستی MRI و CSF برای تشخیص بیماری‌های شناختی بهره برده‌اند. فلگیبل و همکاران (Fellgiebel, Scheurich, Bartenstein, Muller, 2007) از دو نشانگر زیستی PET و CSF برای تشخیص این بیماری‌ها و ولهوود و همکاران (Walhovd et al, 2010) در تحقیقاتشان از هر سه نشانگر زیستی PET، MRI و CSF برای تشخیص بیماری‌های شناختی استفاده کرده‌اند. جانگ (Zhang) و همکاران با استفاده از ویژگی‌های استخراج شده از سه نشانگر زیستی PET، MRI و CSF روشی مبتنی بر یادگیری چند هسته‌ای

1 Multikernel based approach

2 Support Vector Machine(SVM)

ارائه داده‌اند. جانگ و همکاران (Zhang et al, 2011) نشان داده‌اند که نتایج به دست آمده بر اساس ترکیب ویژگی‌های استخراج‌شده از نشانگرهای زیستی متفاوت به مراتب بهتر از نتایج به دست آمده با استفاده از تنها یک نشانگر زیستی است. جائه و همکاران (Biao et al, 2014) با استفاده از تصاویر کارکردی مغز و تشکیل شبکه ارتباطی بین نواحی مشخص‌شده از مغز توانسته‌اند روشی جدید برای تشخیص بیماری‌های شناختی ارائه دهند. آن‌ها با استفاده از روش‌های تحلیل شبکه‌های ارتباطی، ویژگی‌های محلی<sup>۱</sup> و توپولوژیک<sup>۲</sup> شبکه را استخراج و بعد از ساخت ماتریس هسته<sup>۳</sup> هر کدام، آن‌ها را با کمک یادگیری چند هسته‌ای باهم ادغام کرده و به روشی جدید و با دقت بالا برای تشخیص بیماری‌های شناختی رسیده‌اند (Biao et al, 2014).

وی (Wee) و همکاران (Wee, Yap, Zhang, Wang, Shen, 2012) با به‌کارگیری کدگذاری تَنک روی شبکه‌های کارکردی حاصل از تصاویر کارکردی مغز روشی جدید برای تشخیص بیماری آلزایمر ارائه داده‌اند. آن‌ها در این روش برای ساخت شبکه کارکردی مغز از کدگذاری تَنک به منظور حذف ارتباطات غیرمفید استفاده کرده‌اند. چنگ یو وی و همکاران (Wee et al, 2012) نشان داده‌اند که استفاده از کدگذاری تَنک برای ساخت شبکه کارکردی مغز و استفاده از ابزار یادگیری ماشین نهایتاً منجر به بالا رفتن دقت در تشخیص بیماری‌های شناختی می‌شود.

### روش پژوهش

هدف از ارائه این مقاله بالا بردن دقت تشخیص بیماری آلزایمر بر اساس ویژگی‌های استخراج‌شده از نشانگرهای زیستی PET، MRI و CSF است. نکته کلیدی در این مقاله استفاده از کدگذاری تَنک برای حذف ویژگی‌های غیرمفیدی است که از نشانگرهای زیستی فوق استخراج می‌گردند. با استفاده از ویژگی‌های باقیمانده و بر اساس یادگیری چند هسته‌ای ماشین بردار پشتیبان جهت طبقه‌بندی افراد سالم و افراد مبتلا به بیماری آلزایمر آموزش داده می‌شود. نتایج نشان می‌دهد که استفاده از کدگذاری تَنک برای حذف ویژگی‌های غیرمفید موجب بالا رفتن دقت تشخیص این بیماری می‌شود.

### طبقه‌بندی چند هسته‌ای

در این بخش یک مرور کلی روی روش‌های یادگیری مبتنی بر هسته خواهیم داشت. این روش‌ها به راحتی قابل اعمال به ماشین بردار پشتیبان جهت طبقه‌بندی داده‌ها هستند (Scholkopf, Smola, 2002). قبل از پرداختن به روش‌های چند هسته‌ای، ابتدا ماشین بردار تک هسته‌ای را مورد بررسی قرار می‌دهیم. ایده اصلی ماشین بردار پشتیبان به صورت خلاصه بدین شرح است: ابتدا داده‌هایی که به صورت خطی جدا پذیر نیستند از فضای اصلی به فضایی با ابعاد بالاتر<sup>۴</sup> که در آن داده‌ها به صورت خطی جدا پذیر

- 
- 1 Local feature
  - 2 Topological feature
  - 3 Kernel matrix
  - 4 Feature space

باشند نگاشت می شود. این نگاشت از طریق توابع هسته<sup>۱</sup> قابل اعمال است. سپس مدل مورد نظر با بیشینه مقدار حاشیه برای طبقه بندی اطلاعات ساخته می شود.

برای ساخت مدل بر اساس ترکیب ویژگی های حاصل از نشانگرهای زیستی MRI، PET و CSF جهت تشخیص بیماری آلزایمر از ماشین بردار پشتیبان چند هسته ای<sup>۲</sup> استفاده می شود (Zhang et al, 2011). برای بررسی ماشین بردار پشتیبان چند هسته ای فرض می کنیم  $n$  داده آموزشی و  $m$  وجه (به طور مثال نشانگر زیستی) داریم.  $x_i^{(m)}$  نشان دهنده بردار ویژگی  $m$ -امین وجه از  $i$ -امین داده آموزشی است و برچسب معادل آن  $y_i$  یکی از دو مقدار  $-1$  و  $1$  است. ماشین بردار پشتیبان چند هسته ای تابع هدف در رابطه (۱) را بهینه می کند.

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

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

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

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 ویژگی هاست. فرم دوگان ماشین بردار پشتیبان همانند فرم اولیه آن به دنبال بهینه کردن تابع معیار در رابطه (۲) آورده شده است.

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

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

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

در رابطه (۲)،  $k^{(m)}(x_i^{(m)}, x_j^{(m)}) = \Phi^{(m)}(x_i^{(m)})^T \Phi^{(m)}(x_j^{(m)})$ ،  $x_i$  و  $x_j$  روی  $m$ -امین وجه از داده است.

1 Kernel function

2 Multikernel Support Vector Machine (MKLSVM)

برای داده‌ی آزمایشی جدید  $x = \{x^{(1)}, x^{(2)}, \dots, x^{(M)}\}$  ابتدا مقدار  $k^{(m)}(x_i^{(m)}, x_j^{(m)}) = \Phi^{(m)}(x_i^{(m)})^T \Phi^{(m)}(x_j^{(m)})$  که نشان‌دهنده مقدار تابع هسته مورد استفاده میان داده آزمایشی و تمام داده‌های آموزشی روی  $m$ -امین وجه است، محاسبه می‌گردد. سپس رابطه (۳) برای پیش‌بینی برچسب داده آزمایشی اعمال می‌گردد.

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

اگر  $k(x_i, x_j) = \sum_m \beta_m k^{(m)}(x_i^{(m)}, x_j^{(m)})$  را به عنوان تابع هسته‌ی ادغام‌شده بین دو داده‌ی آموزشی  $x_i$  و  $x_j$  و  $k(x_i, x) = \sum_m \beta_m k^{(m)}(x_i^{(m)}, x^{(m)})$  را به عنوان تابع هسته‌ی ادغام‌شده بین داده‌ی آموزشی  $x_i$  و داده‌ی آزمایشی  $x$  در نظر بگیریم، آنگاه می‌توان ماشین بردار پشتیبان چند هسته‌ای را به وسیله‌ی ماشین بردار تک هسته‌ای پیاده‌سازی کرد. در فرمول‌های فوق این محدودیت وجود دارد که  $\sum_m \beta_m = 1$  می‌توان با قرار دادن مقادیر مختلف برای هر  $\beta_m$  در نظر گرفتن محدودیت ذکر شده و با استفاده از اعتبارسنجی  $k$ -تکه‌ای<sup>۱</sup> روی داده‌های آموزشی بهترین مقادیر را برای این متغیرها یافت. در این مقاله برای آموزش بردار پشتیبان از ابزار LIBSVM (Chang, Lin, 2001) استفاده شده است.

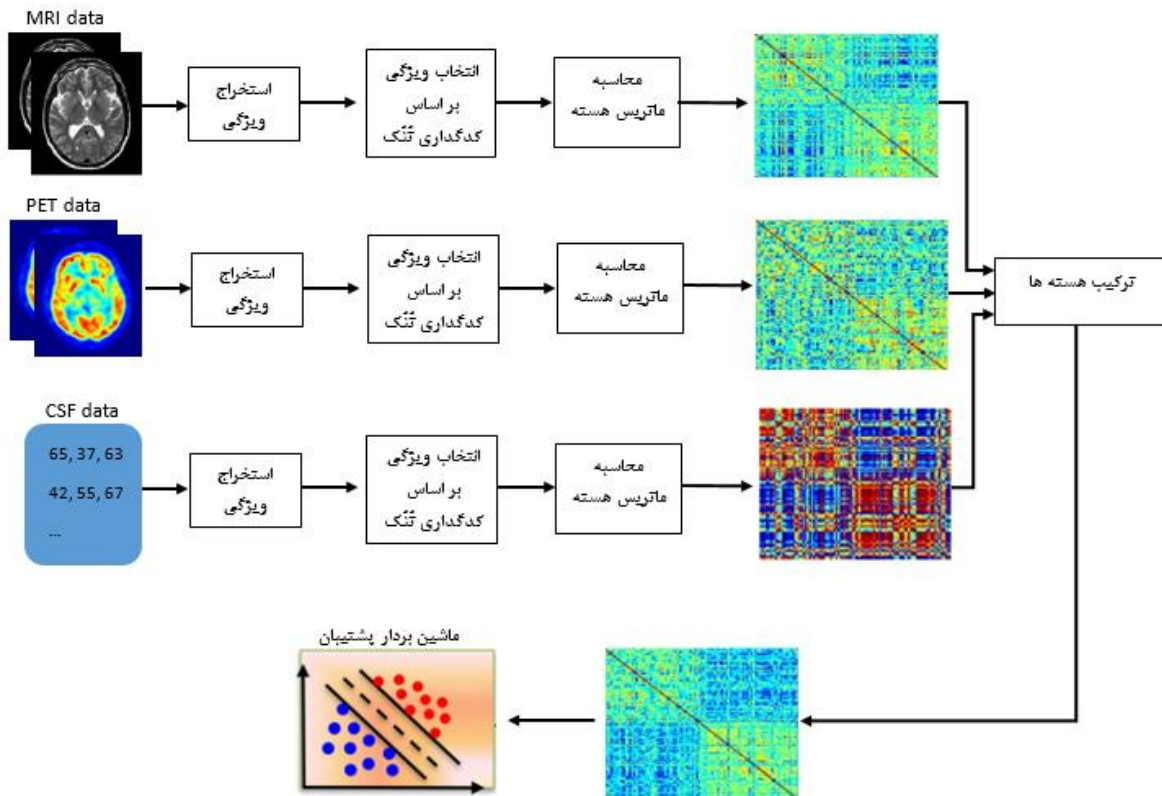
### کدگذاری تُنک

نکته کلیدی در این مقاله استفاده از کدگذاری تُنک جهت حذف ویژگی‌های غیرمفید برای تشخیص بیماری‌های شناختی است. حذف ویژگی‌های غیرمفید با بهینه کردن تابع معیار رابطه (۴) امکان پذیر است.

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

در رابطه (۴)  $x_i$  نشان‌دهنده بردار ویژگی،  $y_i$  نشان‌دهنده برچسبی است که داده‌ی  $x_i$  به آن تعلق دارد،  $b$  مقدار عرض از مبدا داده  $x_i$ ، ضرایب رگرسیون بردار ویژگی داده  $x_i$ ،  $\lambda$  پارامتر تنظیم و  $N$  تعداد داده آموزشی است.  $\|w\|_1$ ، تنظیم‌کننده  $L_1$ -norm است که برای کدگذاری تُنک در فضای ویژگی مورد استفاده قرار می‌گیرد. به عنوان مثال ضرایب رگرسیونی که برای ویژگی‌های غیرمفید با استفاده از این روش صفر می‌شوند. با استفاده از این روش ویژگی‌های معادل با این ضرایب از بردار ویژگی حذف می‌شوند. در این مقاله ویژگی‌های غیرمفید استخراج‌شده از دو نشانگر زیستی MRI و PET برای تشخیص بیماری آلزایمر با استفاده از این روش حذف ویژگی، حذف شده و سپس با استفاده از ویژگی‌های باقیمانده هسته‌های مربوط به هر کدام از این ویژگی‌ها محاسبه و باهم ادغام می‌شوند. در نهایت با استفاده از ماشین بردار پشتیبان و با استفاده از هسته‌های ترکیب‌شده عمل طبقه‌بندی روی افراد سالم و افراد مبتلا به بیماری آلزایمر انجام می‌پذیرد. در شکل (۱) شمای کلی این مراحل به تصویر کشیده شده است. بهترین مقدار برای

پارامتر تنظیم ( $\lambda > 0$ ) در رابطه (۴) با استفاده از عمل اعتبارسنجی  $k$ -تکه‌ای روی داده‌های آموزشی قابل محاسبه است. در این مقاله از ابزار LASSO (Tibshirani, 1996) برای بهینه کردن رابطه (۴) استفاده می‌شود.



شکل(۱): نمایش شماتیک روند کلی کار برای طبقه‌بندی داده‌های چندوجهی<sup>۱</sup>

### اعتبارسنجی

برای سنجیدن کارایی ماشین بردار پشتیبان آموزش داده شده از اعتبارسنجی  $k$ -تکه‌ای به منظور محاسبه دقت طبقه‌بندی استفاده می‌کنیم. طبق این کار مجموعه داده‌های موجود (داده‌های مربوط به ۵۱ فرد مبتلا به آلزایمر و ۵۲ فرد سالم) را به ۱۰ زیرمجموعه افزایش می‌کنیم. در هر بار از ۹ زیرمجموعه برای آموزش ماشین بردار پشتیبان و از یک زیرمجموعه باقیمانده برای سنجش مدل آموزش داده شده استفاده می‌کنیم. این کار به تعداد ۱۰ بار به منظور ساختن زیرمجموعه‌هایی با ترکیب داده‌های مختلف و اطمینان از رخ ندادن Overfitting مدل، تکرار می‌شود. برای یافتن وزن بهینه ماتریس هسته هر کدام از  $m$  نشانگر زیستی ( $\beta_m$ ) به منظور ترکیب



هسته‌ها، از داده‌های آموزشی استفاده می‌کنیم. بدین ترتیب که با قرار دادن مقادیر مختلف از ۰ تا ۱ و اندازه گام ۰,۱ برای  $\beta_m$ ها و آموزش مدل بر اساس این مقادیر (با در نظر گرفتن محدودیت  $\sum_m \beta_m = 1$ )، دقت مدل را بر حسب داده‌های آزمایشی می‌سنجیم و وزن‌هایی که موجب بیش‌ترین دقت برای مدل شده‌اند را به عنوان وزن هسته‌ها در نظر می‌گیریم. برای نرمال‌سازی داده‌ها به صورت رابطه (۵) عمل می‌کنیم.

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

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

### خلاصه روش پیشنهادی

در شکل (۱) روند کلی کار در روش پیشنهادی به تصویر کشیده شده است، همان‌طور که در این شکل می‌بینید ابتدا ویژگی‌های مورد نیاز از سه نشانگر زیستی MRI، PET و CSF استخراج می‌گردند. تابع معیار نشان داده‌شده در رابطه (۴) با استفاده از ابزار LASSO به منظور انجام عمل کدگذاری تُنک جهت حذف ویژگی‌های غیرمفید بهینه می‌گردد. در مرحله بعد ماتریس هسته مربوط به هر کدام از نشانگرهای زیستی فوق با استفاده از ویژگی‌های باقیمانده محاسبه می‌گردد. با یافتن بهترین اوزان ماتریس‌های هسته به دست آمده، این ماتریس‌ها باهم ترکیب‌شده و ماشین بردار پشتیبان جهت انجام عمل طبقه‌بندی آموزش داده می‌شود.

در شکل (۲) شبه کد مربوط به روش پیشنهادی آورده شده است. ورودی تابع تعریف‌شده ویژگی‌های استخراج‌شده از نشانگرهای زیستی MRI، PET و CSF و خروجی آن دقت ماشین بردار پشتیبان آموزش داده‌شده جهت تشخیص بیماری آلزایمر است.

#### Pseudocode for the proposed method

```

Function AD_HC_ACC (biomarkers features)
  Select useful features with sparse coding
  Compute kernel matrixes
  Compute weight of each kernel matrix
  Combine all matrixes
  Train SVM
  Compute accuracy with k-fold cross validation
End function
    
```

شکل (۲): شبه کد مربوط به روش پیشنهادی

### ارزیابی

در این بخش به بررسی داده‌های مورد استفاده در این مقاله و نتایج به دست آمده در پیاده‌سازی روش پیشنهادی می‌پردازیم. برای ارزیابی نتایج به دست آمده با روش جدید، آن‌ها را با نتایج روش‌های رایج برای طبقه‌بندی مانند شبکه عصبی، درخت تصمیم و ... مقایسه می‌کنیم.

### تجزیه و تحلیل داده‌ها

داده‌های مورد استفاده در این مقاله از پایگاه‌های داده آزمایشگاه تصویربرداری عصبی بیماری آلزایمر<sup>۱</sup> (ADNI) استخراج شده است.<sup>۲</sup> در این آزمایشگاه از مغز افراد مختلف تصویربرداری می‌شود. افرادی که در این آزمایشگاه مورد تصویربرداری عصبی قرار می‌گیرند دارای سنی بین ۵۵ و ۹۰ سال بوده و عبارت‌اند از ۱- افراد سالم<sup>۳</sup>، ۲- افراد مبتلا به اختلالات خفیف شناختی و ۳- افراد مبتلا به بیماری آلزایمر. در این مقاله از اطلاعات مربوط به ۱۰۳ فرد مورد آزمایش در ADNI که هر سه نشانگر زیستی MRI، PET و CSF برای آن‌ها محاسبه شده، استفاده می‌کنیم. این ۱۰۳ نفر عبارت‌اند از: ۵۱ فرد مبتلا به بیماری آلزایمر و ۵۲ فرد سالم. برای استخراج ویژگی‌های مربوط به نشانگرهای زیستی حاصل از MRI و PET تصاویر مغز حاصل از تصویربرداری ساختاری و کارکردی به ۹۳ ناحیه<sup>۴</sup> تقسیم می‌شود. 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 پیش‌پردازش‌هایی همچون حذف حرکات اضافی سر<sup>۵</sup>، اصلاح زمانی سری زمانی تصاویر<sup>۶</sup> برای هر فرد، و اصلاح مکانی<sup>۷</sup> هر کدام از تصاویر موجود در سری زمانی نسبت به اولین تصویر موجود در این سری زمانی برای هر فرد، مقدار میانگین شدت هر ناحیه در مغز به عنوان ویژگی آن در نظر گرفته می‌شود. برای نشانگر زیستی CSF نیز سه مقدار CSF Aβ42، CSF t-tau، CSF p-tau استفاده شده است (Zhang et al, 2011). بنابراین برای هر فرد در مجموع ۹۳ ویژگی از تصویر MRI، ۹۳ ویژگی از تصویر PET و ۳ ویژگی از مایع مغزی نخاعی در نظر گرفته می‌شود.

### نتایج

همان طور که در جدول (۱) نشان داده شده است، عمل طبقه‌بندی را روی داده‌های AD (۵۱ فرد مبتلا به بیماری آلزایمر) و HC (۵۲ سالم) انجام داده‌ایم. برای انجام مقایسه و سنجیدن دقت روش پیشنهادی، روش‌های درخت تصمیم<sup>۸</sup>، شبکه عصبی<sup>۹</sup> و k-نزدیک‌ترین همسایه<sup>۱۰</sup> و ماشین بردار پشتیبان را نیز به منظور طبقه‌بندی داده‌های مورد نظر اجرا کرده‌ایم که نتایج در جدول (۱) نشان داده شده است. روش پیشنهادی در محیط MATLAB پیاده‌سازی و نتایج آزمایش روی میانگین و خطای استاندارد ۱۰ بار آموزش و آزمایش مدل گزارش شده است. آموزش مدل با استفاده از اعتبارسنجی ۱۰-تکه‌ای انجام شده است، به این صورت که داده‌های موجود به ۱۰ قسمت تقسیم شده و در هر مرحله ۹ قسمت از داده‌ها برای آموزش مدل و ۱ قسمت باقیمانده برای سنجیدن آن به کار می‌رود. این فرآیند برای هر روش آزمایش (KNN، NN و ...) به تعداد ۱۰ بار اجرا شده و دقت مدل بر اساس میانگین و با در نظر گرفتن خطای استاندارد<sup>۱۱</sup> در جدول (۱) گزارش شده است. سطرهای ۱ تا ۳ این جدول نتایج به دست آمده از

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

8 Decision Tree (DT)

9 Neural Network (NN)

10 K\_nearest neighbor (KNN)

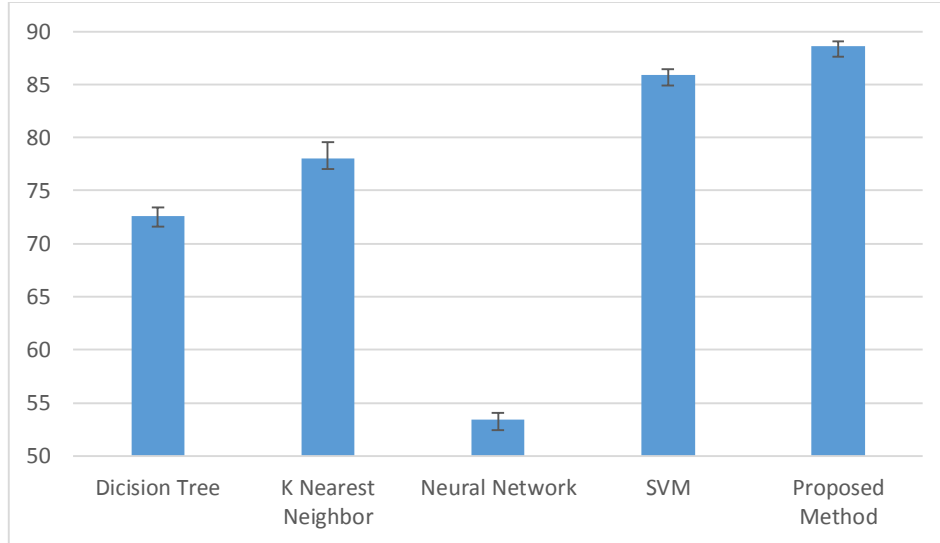
11 Standard error

اجرای روش‌های گفته‌شده با استفاده از هر یک از نشان‌گرهای زیستی، در سطرهای چهارم تا ششم، نتایج با استفاده از ترکیب‌های دوتایی مختلف 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 در نظر گرفته شده است.

جدول ۱- نتایج حاصل از اجرای روش‌های طبقه‌بندی روی داده‌ها بر حسب میانگین دقت  $\pm$  خطای استاندارد در ۱۰ بار اجرای مستقل

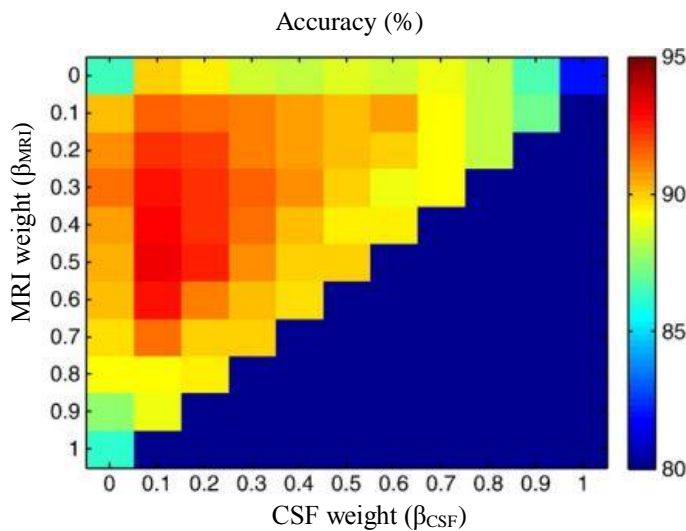
روش پیشنهادی	ماشین بردار پشتیبان	K-نزدیک‌ترین همسایه	شبکه عصبی	درخت تصمیم	نشان گر زیستی
۸۶/۹۱ $\pm$ ۰/۵۵	۸۰/۵۷ $\pm$ ۰/۷۱	۷۸/۴۱ $\pm$ ۰/۳۱	۵۲/۶۵ $\pm$ ۱/۲۵	۷۲/۹۵ $\pm$ ۰/۶۵	<b>MRI</b>
۸۹/۷۲ $\pm$ ۰/۳۶	۸۶/۴۱ $\pm$ ۰/۵۹	۷۶/۷۸ $\pm$ ۰/۵۷	۵۱/۷ $\pm$ ۱/۱۸	۶۸/۴۱ $\pm$ ۰/۸۵	<b>PET</b>
۸۰/۱۴ $\pm$ ۰/۶۱	۷۹/۷ $\pm$ ۰/۴۲	۷۹/۳۷ $\pm$ ۰/۳۶	۵۴/۰۳ $\pm$ ۱/۷۲	۶۷/۸ $\pm$ ۰/۹۴	<b>CSF</b>
۹۰/۲۶ $\pm$ ۰/۳	۸۶/۸۱ $\pm$ ۰/۲۶	۷۸/۱ $\pm$ ۰/۶۳	۵۵/۲۰ $\pm$ ۱/۷۷	۷۳/۵۹ $\pm$ ۰/۹	<b>MRI+PET</b>
۸۶/۸ $\pm$ ۰/۶	۸۳/۷ $\pm$ ۰/۵۶	۸۰/۰۶ $\pm$ ۰/۸۵	۵۵/۳۳ $\pm$ ۱/۷۳	۷۱/۵ $\pm$ ۰/۶۶	<b>MRI+CSF</b>
۹۲/۶۶ $\pm$ ۰/۴۲	۸۹/۰۷ $\pm$ ۰/۷۸	۷۵/۶۶ $\pm$ ۰/۹۷	۵۴/۴۱ $\pm$ ۱/۷۸	۷۶/۱ $\pm$ ۰/۸۸	<b>CSF+PET</b>
۹۳/۹ $\pm$ ۰/۴۱	۸۹/۱۱ $\pm$ ۰/۴۷	۷۸/۰۶ $\pm$ ۱/۰۷	۵۴/۶۵ $\pm$ ۱/۲۲	۷۷/۶۲ $\pm$ ۰/۷۵	<b>MRI+PET+CSF</b>

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 ویژگی‌های یک نشان گر زیستی استفاده می‌شود، ابتدا با استفاده از کد گذاری تُنک ویژگی‌های غیرمفید حذف شده و در نهایت ماشین بردار پشتیبان را آموزش می‌دهیم. همچنین در حالتی که از ترکیب‌های مختلف این نشانگرهای زیستی استفاده می‌شود، ابتدا ویژگی‌های 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 ماتریس‌ها آموزش داده می‌شود. نتایج نشان می‌دهند که روش پیشنهادی موجب بهبود دقت تشخیص بیماری آلزایمر می‌گردد. در شکل (۳) نمودار مربوط به میانگین دقت و خطای استاندارد به دست آمده در پیاده‌سازی روش‌های گفته‌شده در جدول (۱) آورده شده است. این نمودار نشان می‌دهد که روش پیشنهادی دارای بیشترین دقت و کمترین خطای استاندارد است.



شکل (۳): نمودار درصد صحت و خطای استاندارد روش های مختلف

در شکل (۴) دقت ماشین بردار آموزش داده شده بر اساس مقادیر مختلف برای وزن هر کدام از نشانگرهای زیستی یعنی  $\beta_{MRI}$ ،  $\beta_{PET}$  و  $\beta_{CSF}$  قابل مشاهده است. در این شکل دو مقدار  $\beta_{CSF}$  و  $\beta_{MRI}$  آورده شده است. طبق این محدودیت که  $\sum_m \beta_m = 1$  بنابراین  $\beta_{PET} = 1 - (\beta_{MRI} + \beta_{CSF})$  مشاهده می شود که مقادیری در حدود  $\beta_{MRI} = 0.5$ ،  $\beta_{PET} = 0.4$  و  $\beta_{CSF} = 0.1$  بهترین دقت را جهت طبقه بندی افراد مبتلا به آلزایمر و افراد سالم فراهم می آورند.



شکل (۴): دقت ماشین بردار پشتیبان آموزش داده شده بر اساس وزن های مختلف برای نشانگرهای زیستی

### نتیجه‌گیری

در این مقاله طرحی جدید مبنی بر استفاده از کدگذاری تُنک به منظور حذف ویژگی‌های غیرمفید استخراج‌شده از نشانگرهای زیستی، برای بالا بردن دقت در تشخیص بیماری‌های شناختی مانند آلزایمر ارائه شده است. این روش موجب بالا رفتن دقت تشخیص این بیماری‌ها نسبت به روش‌های قبلی می‌شود. در روش اولیه ویژگی‌های استخراج‌شده از نشانگرهای زیستی مختلف به صورت یکجا در نظر گرفته شده و عمل طبقه‌بندی روی داده‌ها به کمک این ویژگی‌ها انجام می‌گیرد. در روشی دیگر با استفاده از روش‌های یادگیری چند هسته‌ای، برای هر کدام از نشانگرهای زیستی هسته‌های جداگانه در نظر گرفته شده و با استفاده از ادغام این هسته‌ها عمل طبقه‌بندی صورت می‌پذیرد. نتایج به دست آمده با استفاده از داده‌های مربوط به ۵۱ فرد مبتلا به بیماری آلزایمر و ۵۲ فرد سالم نشان می‌دهد که روش پیشنهادی این مقاله مبنی بر حذف ویژگی‌های غیرمفید برای تشخیص بیماری آلزایمر با استفاده از کدگذاری تُنک و استفاده از ویژگی‌های باقیمانده برای عمل طبقه‌بندی موجب بالا رفتن دقت تشخیص این بیماری می‌شود.

تاکنون مطالعات بسیار زیادی برای تشخیص بیماری‌های شناختی انجام شده و روش‌های بسیار متنوعی برای بالا بردن قدرت تشخیص این بیماری‌ها ارائه شده است. از جمله این روش‌ها ساخت شبکه کارکردی از روی تصاویر کارکردی مغز جهت تشخیص بیماری‌های شناختی است. در کارهای آتی به بررسی روش‌های مختلف استخراج ویژگی با استفاده از مفاهیم موجود در تحلیل شبکه‌های ارتباطی جهت استخراج ویژگی‌های موثر از شبکه‌های کارکردی مغز در تشخیص بیماری‌های 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.

- 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-tECTION 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.
- Lanckriet 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., Pirtila, 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. Royal Statistical Soc. Series B Methodol.*, vol. 58, pp. 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.