

# 뇌기능영상기반 기능적 연결성 행렬의 서로 다른 성분 비율과 주성분 벡터에 따른 자폐 스펙트럼 장애의 랜덤 포레스트 분류성능 비교

최형신<sup>1</sup>, 박현진<sup>2,3\*</sup>

<sup>1</sup>성균관대학교 전기전자컴퓨터공학과

<sup>2</sup>성균관대학교 전자전기공학부

<sup>3</sup>기초과학연구원 뇌과학이미징연구단

[gudt1s17@naver.com](mailto:gudt1s17@naver.com), [hyunjinp@skku.edu](mailto:hyunjinp@skku.edu)\*

## Comparison of random forest classification performance of autism spectrum disorders according to different component ratios of the functional connectivity matrix and principal component vectors using neuroimaging

Hyounghsin Choi<sup>1</sup>, Hyunjin Park<sup>2,3\*</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea

<sup>2</sup>Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Korea

<sup>3</sup>School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, Korea

### 요 약

자폐 스펙트럼 장애는 이질적인 신경 발달 장애로, 뇌기능영상에 기반한 기능적 연결성 행렬을 이용해 연구가 활발하게 진행된다. 기능적 연결성 행렬을 분석하기 위해 주성분 분석방법을 이용하며, 이를 통해 뇌의 기능적 경향성 패턴을 확인할 수 있다. 이 때, 서로 다른 연결성 성분 비율과 주성분 벡터를 이용해서 다양한 기능적 경향성 패턴을 얻을 수 있다. 패턴에 따른 랜덤 포레스트 분류 모델의 성능이 달라지는데 이를 비교해본 결과, 상위 50%의 성분을 이용하여 만든 기능적 경향성 패턴 1이 데이터의 설명 비율도 높고, 우수한 분류 성능을 보이는 것을 확인했다.

### 1. 서론

자폐 스펙트럼 장애 (Autism spectrum disorder, ASD)는 이질적인 신경 발달 장애다. 사회적 의사소통에 어려움이 있고, 제한적이고 반복적인 행동이 특징이다[1]. 기존 연구들은 ASD를 연구하기 위해 뇌기능영상에 기반한 기능적 연결성 행렬 (Functional connectivity, FC)을 사용했다[1]. FC는 뇌 영역들의 활성화가 통계적으로 서로 연관되어 있는지 계산하는 방법으로[2], 각 피험자들이 속한 집단의 요인과 피험자들이 갖는 개별 요인으로 구성되어 있다. 이러한 요인들로 인해 ASD 피험자와 정상 발달 (Typical development, TD) 피험자 사이의 뇌 활성화를 비교할 수 있다. 기존 연구들에 의하면, 고차원 데이터인 FC를 분석하기 위해 FC 데이터를 가공한다[3]. FC 데이터를 가공하기 위해, 강한 연결성을 갖는 성분만 남긴 후, 차원 축소 기법 중 하나인 주성분 분석 (Principal

component analysis, PCA)을 적용하여 분석을 진행한다. 이 때 다양한 주성분을 얻을 수 있고, 이를 이용해서 뇌의 기능적 경향성 패턴을 확인할 수 있다. 그리고 기능적 경향성 패턴의 차이를 이용해서 ASD 집단과 TD 집단을 구분할 수 있다. 따라서 본 연구에서는 ASD 집단과 TD 집단을 잘 구분하는 방법을 찾기 위해 다양한 가공법을 적용해본 후 랜덤포레스트 분류 성능을 비교해보려 한다.

### 2. 방법

본 연구는 Autism Brain Imaging Data exchange-I (ABIDE-I) 데이터베이스로부터 ASD 피험자 103명, TD 피험자 108명, 총 211명 피험자의 휴지 상태 기능성 자기공명영상 (resting state functional MRI, rs-fMRI) 데이터를 얻어서 진행되었다[4]. ASD는 성별에 영향을 받기 때문에, 이번 연구에서는 남성 데이터만 사용하였다[5]. 뇌 영역은 기능과 구조에 따라 표준

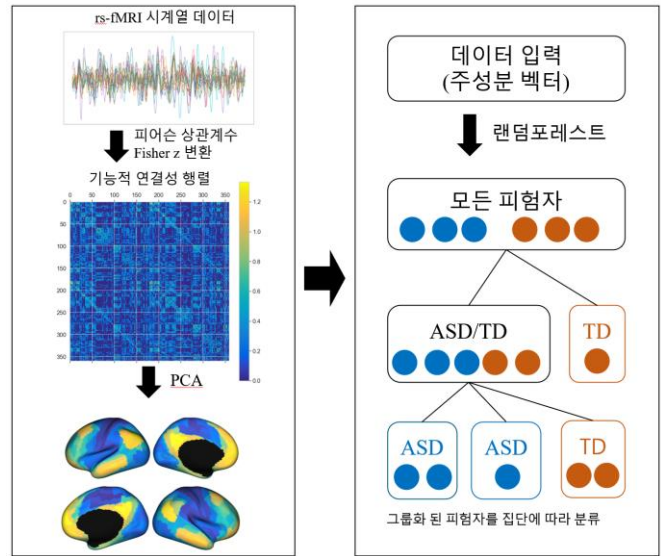
공간에서 360 개의 영역으로 나타낸 표준뇌지도인 multimodal parcellation (MMP)[6]에서 정의된 뇌 영역의 서로 다른 두 영역의 rs-fMRI 시계열 데이터의 피어슨 상관계수를 이용하여 FC 를 계산했다. 그리고, 계산된 FC 행렬에 Fisher z 변환을 적용한 후, 각 행별로 강한 연결성을 갖는 성분만 남겼다. ASD 집단과 TD 집단의 분류에 끼치는 영향력을 확인하기 위해 각각 상위 10%, 25%, 50%의 성분을 남기고 연구를 진행했다.

PCA 는 고차원 데이터를 저차원으로 변환하는 선형 변환 방법이다. 최대 분산을 설명하는 직교 성분 집합으로 변환시킨다[3]. 각각의 주성분 고유 벡터는 뇌의 활성화 패턴을 나타내기 때문에, 강한 연결성을 갖는 성분만 남긴 FC 행렬의 고윳값을 분해해서 행렬 성분들의 최대 분산을 설명하는 주성분 고유 벡터들을 구하였다. 1,2,3 주성분 벡터를 구해서 집단 분류에 끼치는 영향력을 확인했다. 각각의 주성분 벡터를 같은 공간에서 비교하기 위해, 피험자 별로 계산된 주성분 벡터를 집단 평균 FC 행렬의 주성분 벡터로 Procrustes 정렬을 하였다.

각각의 주성분 벡터를 통해 구한 기능적 경향성 패턴에서 ASD/TD 집단 별 유의한 차이가 나는 뇌 영역을 선택하기 위해 least absolute shrinkage and selection operator (LASSO) 변수 선택 방법을 이용했다.

### 3. 결과

서로 다른 성분 비율을 갖는 FC 를 이용하여 주성분 벡터 1,2,3 을 구했다. 각각의 주성분 벡터에 따른 3 개의 기능적 경향성 패턴을 확인하고, 각각의 패턴에서 유의한 뇌 영역을 이용하여 랜덤포레스트 분류를 한 결과, <표 1>과 같은 결과를 확인할 수 있었다. FC 의 상위 10%의 성분을 이용해서 만든 기능적 경

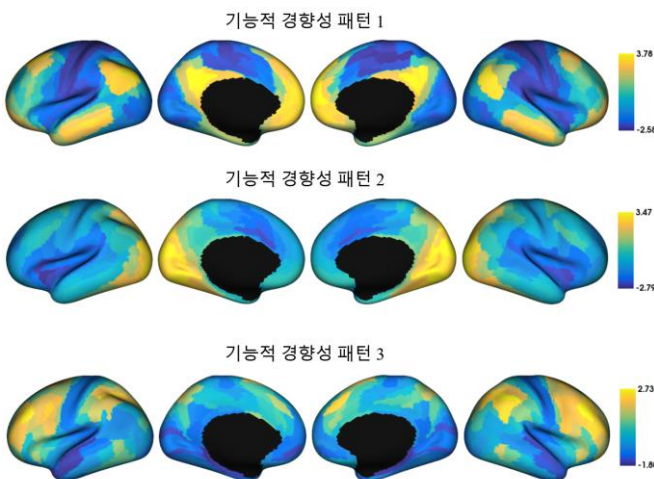


(그림 2) 전체적인 연구방법.

향성 패턴 3 가 72.8%로 가장 높은 분류 성능을 보였지만, 데이터의 11.3%만 설명하는 것으로 나타났다. 그리고 모든 성분을 사용할 때, 그리고 상위 50%의 성분을 사용해서 만든 기능적 경향성 패턴 1 이 39.4%, 39.3%로 다른 성분 비율에 비해 데이터를 가장 많이 설명하는 것으로 나타났다. 특히, 상위 50%의 성분을 이용해서 만든 패턴 1 의 경우, 65.1%의 분류 성능을 나타냈다.

### 4. 결론

이번 연구를 통해 뇌기능영상에서 얻은 뇌의 기능적 경향성 패턴이 ASD 집단과 TD 집단을 구별할 수 있는 특성을 갖는 것을 확인할 수 있었다. 각각의 기능적 경향성 패턴이 서로 다른 분류 성능을 보였고, 데이터의 설명 비율도 다르기 때문에, 가지고 있는 특성도 다를 것으로 보여진다. 향후 연구에서 각 패턴이 설명하는 특성을 파악할 수 있다면 분류 성능 개선에 많은 도움이 될 것으로 보여진다.



(그림 1) 주성분 벡터로 구한 기능적 경향성 패턴.

FC 의 성분 비율과 주성분 벡터를 다르게 해서 구한 기능적 경향성 패턴에서 집단 별 유의한 차이가 나는 뇌 영역을 구한 후, 이를 이용해서 랜덤포레스트 분류모델을 만들었다. 5-fold 교차검증을 해서 평균 정확도 민감도, 특이도를 측정해서 모델의 성능을 평가하였다.

<표 1> 남겨진 성분 비율과 사용한 기능적 경향성 패턴에 따른 모델 분류 성능

성분비율	정확도	민감도	특이도	분산설명비율
<b>모든 성분</b>				
기능적 경향성 패턴 1	63.0 ± 4.03	61.3 ± 7.64	64.6 ± 7.52	<b>39.4</b>
기능적 경향성 패턴 3	53.7 ± 4.30	49.7 ± 7.76	57.6 ± 8.13	25.9
기능적 경향성 패턴 3	67.4 ± 4.60	62.0 ± 8.51	72.6 ± 7.26	12.8
<b>상위 10%</b>				
기능적 경향성 패턴 1	55.4 ± 4.43	53.8 ± 7.97	56.9 ± 7.73	27.9
기능적 경향성 패턴 3	60.2 ± 4.31	53.0 ± 7.88	67.0 ± 7.60	20.4
기능적 경향성 패턴 3	<b>72.8 ± 4.21</b>	<b>67.3 ± 7.45</b>	<b>78.0 ± 6.44</b>	11.3
<b>상위 25%</b>				
기능적 경향성 패턴 1	62.0 ± 4.86	58.1 ± 8.58	65.7 ± 7.79	34.9
기능적 경향성 패턴 3	58.2 ± 4.21	57.2 ± 7.97	59.3 ± 7.46	29.9
기능적 경향성 패턴 3	61.0 ± 4.11	61.0 ± 7.82	61.0 ± 7.18	10.0
<b>상위 50%</b>				
기능적 경향성 패턴 1	<b>65.1 ± 4.06</b>	<b>62.9 ± 7.52</b>	<b>67.2 ± 7.43</b>	<b>39.3</b>
기능적 경향성 패턴 3	51.4 ± 4.29	52.4 ± 8.27	50.4 ± 7.39	28.6
기능적 경향성 패턴 3	62.6 ± 4.22	62.5 ± 7.95	62.6 ± 7.30	12.6

## 5. Acknowledgements

This research was supported by National Research Foundation (NRF-2020M3E5D2A01084892), Institute for Basic Science (IBS-R015-D1), Ministry of Science and ICT (IITP-2020-2018-0-01798), IITP grant funded by the AI Graduate School Support Program (2019-0-00421), and ICT Creative Consilience program (IITP-2020-0-01821).

## 참고문헌

- [1] M. A. Just, V. L. Cherkassky, T. A. Keller, and N. J. Minschew, "Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity," *Brain*, 2004, doi: 10.1093/brain/awh199.
- [2] A. K. Easson, Z. Fatima, and A. R. McIntosh, "Functional connectivity-based subtypes of individuals with and without autism spectrum disorder," *Netw. Neurosci.*, 2019, doi: 10.1162/netn\_a\_00067.
- [3] D. S. Margulies *et al.*, "Situating the default-mode network along a principal gradient of macroscale cortical organization," *Proc. Natl. Acad. Sci. U. S. A.*, 2016, doi: 10.1073/pnas.1608282113.
- [4] C. Cameron *et al.*, "The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives," *Front. Neuroinform.*, 2013, doi: 10.3389/conf.fninf.2013.09.00041.
- [5] S. J. Hong *et al.*, "Atypical functional connectome hierarchy in autism," *Nat. Commun.*, 2019, doi: 10.1038/s41467-019-08944-1.
- [6] M. F. Glasser *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, 2016, doi: 10.1038/nature18933.