



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

뇌 발달 질환 빅 코호트 데이터
이용에 있어서 딥 뉴럴 네트워크의
최적화

연세대학교 대학원
의 과 학 과
방 현 석

뇌 발달 질환 빅 코호트 데이터
이용에 있어서 딥 뉴럴 네트워크의
최적화

연세대학교 대학원
의 과 학 과
방 현 석

뇌 발달 질환 빅 코호트 데이터 이용에 있어서 딥 뉴럴 네트워크의 최적화

지도교수 박 해 정

이 논문을 석사 학위논문으로 제출함

2019 년 12 월

연세대학교 대학원

의 과 학 과

방 현 석

방현석 의 석사 학위논문을 인준함

심사위원 박 해 정 인

심사위원 전 용 호 인

심사위원 정 호 성 인

연세대학교 대학원

2019 년 12 월

감사의 글

내용

뇌 과학에 대해 아무것도 모르고 시작한 석사 과정을 진행하는 동안 정말 많은 것을 배웠습니다. 아직 많이 부족하고 기대한 결과를 만들어 내지는 못했지만 지난 2년을 되돌아보면 연구실을 다니고 수업을 들으며 많은 것을 배우고 발전했다는 것을 느낍니다. 이 많은 발전을 하게 도와주신 박해정 지도교수님께 큰 감사를 드립니다. 뇌 과학에 대한 기본적인 지식과, 연구해 나가야 할 방향 뿐 아니라 저의 개인적인 생활과 마음의 단단함을 만들어 주셔서 정말 감사드립니다. 항상 연구실에서 제가 모르는 것을 알려주시고 연구에 대해 매일같이 이야기를 주고받던 연구실의 모든 분께 감사드립니다. 강지영 선생님, 배종원 선생님, 손준호 선생님, 어진석 선생님, 유솔 선생님, 이동하 선생님, 장창원 선생님, 정계삼 선생님, 조수원 선생님, 진승현 선생님, 최윤경 선생님, 최혜윤 선생님께 큰 감사말씀 드립니다. 지도교수님을 포함한 모든 선생님들, 앞으로 훌륭한 연구하시어 인류에 도움이 되는 연구결과를 만들어 내시길 진심으로 바라겠습니다.

저자 씀

차례

국문요약	8
I. 서론	9
1. 연구 배경	9
2. 연구 목적	12
3. 연구 범위	13
II. 재료 및 방법	14
1. 데이터 수집	14
2. 데이터 전처리	15
3. 기능적 연결성 변화 계산	16
4. 중요 연결성 추출	17
5. 분류 모델 구축과 학습	18
가. Graph Convolution Neural Network	18
나. Support Vector Machine	21
다. Artificial Neural Network	22
III. 결과	24
1. 데이터 전 처리와 오류 검출	24
2. 기관별 파라미터 보정 결과	25
3. 중요 연결성 추출 결과	27
4. FC & dFC 분류 성능 비교	28
5. 모델 별 분류 성능 비교	30
IV. 고찰	31
V. 결론	35
참고문헌	37
ABSTRACT	40

그림 차례

그림 1. Graph CNN 구조	19
그림 2. 인공 뉴럴 네트워크 구조	23
그림 3. 데이터 전 처리와 오류 검출.....	24
그림 4. 파라미터 보정 전 후 연결성의 차이 비교..	25
그림 5. 파라미터 보정 전 후 분류학습 성능 비교..	26
그림 6. 중요 연결성 추출.....	27
그림 7. FC & dFC 분류 성능 비교	28
그림 8. 분류 모델 별 성능 비교.....	30

표 차례

표 1. 기관 별 파라미터 정보	15
-------------------------	----

국문요약

뇌 발달 질환 빅 코호트 데이터 이용에 있어서 딥 뉴럴 네트워크의 최적화

기존의 설문 응답을 기반으로 한 주의력 결핍 및 과잉 행동 장애(ADHD)나 자폐증(ASD)등의 아동 발달 장애의 진단은 신뢰도나 응답 성실도에 대한 문제가 제기되어 왔고, 기능적 연결성 등의 바이오 마커 들로 진단하는 연구들이 지속되어왔다.

하지만 머신 러닝을 통한 질환 진단에 있어서, 과거에는 모델 학습을 위한 샘플이 부족하다는 한계를 가지고 있었고, 한 진단의 세부 진단이나 진단 간의 분류를 하는 연구가 부족한 상황이다. 정상 군과 질환 군 두 그룹의 진단이 아닌 여러 그룹을 한번에 진단할 수 있는 기준을 만드는 것이 한 질환의 특징을 확인하는데에 중요한 방법이기 때문에 이러한 다중 분류는 반드시 연구될 필요가 있다. 기존 연구의 한계를 벗어나기 위해 두 가지의 연구를 진행했고, 각각 병원에서 다른 파라미터로 얻어진 데이터에 대한 전처리 보정과, 우수한 머신 러닝 분류 모델을 구축했다. 또한 아동 발달 장애의 행동적 특성을 고려하여 휴지기 기능적 연결성의 변화를 계산하고 그것으로 모델 학습 및 분류를 진행하여 기능적 연결성으로 학습한 모델과 성능 비교 후, 기능적 연결성의 변화가 질환 간 차이를 더 잘 보여준다는 것을 확인했다. 또한 이 과정에서 영역 간 기능적 연결성으로 연결되어 있는 그래프 구조의 특성을 가지고 심층 인공 신경망을 사용하여 전통적인 분류 모델인 서포트 벡터 머신과의 성능 비교를 진행하였다.

핵심되는 말 : 아동 발달장애, Graph CNN, Support Vector Machine

뇌 발달 질환 빅 코호트 데이터 이용에 있어서 딥 뉴럴 네트워크의 최적화

<지도교수 박해정>

연세대학교 대학원 의과학과

방 현 석

I. 서론

1. 연구 배경

기존의 주의력 결핍 및 과잉 행동 장애(ADHD)나 자폐증(ASD)등의 아동 발달 장애의 진단은 임상 심리 분야에서 환자나 보호자의 설문에 대한 응답을 기반으로 하여 진단되어왔다. 하지만 이런 종류의 진단에는 두 가지의 큰 문제점이 존재한다. 첫 번째로 설문 응답에 대한 신뢰도의 문제이다. 대표적인 아동 발달 장애 진단의 지표가 되는 설문 중 아동 자가 진단(YSR) 설문은 100여개의 설문으로 되어있어 아동 발달 장애를 가지고 있을 수 있는 아동이 설문을 완수하기에는 설문의 개수가 너무 많고 그 설문을 끝까지 집중해서 응답했다고 생각하기 쉽지 않다. 또한 아동 본인이 스스로 생각하는 자신의 상태와, 보호자가 생각하는 아동의 상태에 대한 차이가 존재할 수 밖에 없기 때문에 이 차이에 대한 진단의 차이가 생길

수 있다. 두 번째 문제로는 아동 발달 장애 진단에 대한 ‘Gold Standard’가 없다는 것이다. 정상 군과 질환 군을 구분하기 위한 정확한 지표가 있어야 하는데 정상 군과 질환 군의 경계에 있는 사람들을 정확히 어느 한 쪽으로 진단하기 어렵고, 진단을 확실히 내려줄 수 있는 지표가 부족하다. 그렇기 때문에 지금까지 기능적 연결성과 같은 바이오 마커 들을 이용한 연구가 이루어져왔다.¹

주된 연구로 아동 발달 장애가 정상 군에 비해 뇌의 어떤 부분의 기능적 연결성이 차이가 나는지 밝히는 연구가 있었다. 뇌의 세부 네트워크나 특정 영역으로 제한된 부분에서 차이를 밝히고 기능적 연결성의 증가나 감소를 찾아내는 연구가 진행되어왔다.^{2 3 4 5}

또 기능적 연결성은 크게 휴지기 상태와 과제를 하는 상태에서의 연구인 2가지로 나눌 수 있는데, 휴지기 상태 뿐 아니라 어떤 행동 과제를 수행하는 동안의 기능적 연결성을 계산하고 그에 따른 차이를 보는 연구도 진행되어왔다.^{6 7}

기존 연구에서는 휴지기 fMRI(rs-fMRI)를 사용한 아동 발달장애 진단에 대한 연구가 지속적으로 연구되어 왔고, 휴지기 fMRI 영상을 통해 얻어진 기능적 연결성 가지고 머신 러닝을 사용해 정상 군과 질환 군을 분류하는 연구가 진행되고 있다.⁸ 아동 발달 장애 질환 중 주의력 결핍 및 과잉 행동 장애와 자폐증에 대한 연구가 지속적으로 진행되어 왔다.

자폐증과 정상 군의 분류 학습 연구에서는 기존 전통적인 분류 학습인 서포트 벡터 머신과 랜덤 포레스트 에서 나아가 뉴럴 네트워크 기반의 모델인 컨볼루션 뉴럴 네트워크와 오토 인코더를 사용하여 200개의

관심영역으로 나뉜 아틀라스를 사용한 휴지기 기능적 연결성을 학습하여 각각 70%, 80%의 분류 성능을 얻어내었다.^{9 10}

반면 주의력 결핍 및 과잉 행동 장애의 연구에서는 휴지기 기능적 연결성을 사용하긴 했지만, 구조 영상을 사용하여 그룹간의 차이가 나는 영역을 선택한 뒤 기능적 연결성과 조합을 하여 분류 학습을 진행한 연구가 많았다. 이 연구들은 적게는 160개의 관심 영역부터 많게는 400개의 영역으로 이루어진 아틀라스를 사용하여 구조적 차이 및 휴지기 기능적 연결성을 계산하였고, 대부분의 연구에서 전통적인 분류 학습 모델인 서포트 벡터 머신을 사용하였다. 이 연구들의 분류 성능 결과는 최소 73%부터 최고 92%까지의 결과를 보이고 있다.^{11, 12}

휴지기 fMRI의 장점은 5분 내외의 촬영으로 설문 응답보다 더 짧은 시간 동안 검사를 진행할 수 있고, 본인이 스스로 응답하는 것이 아닌, 뇌의 활동에 의한 혈중 산소 포화도 신호를 측정하는 것으로써, 좀더 객관적인 데이터이다. 이런 장점을 가지고 휴지기 기능적 연결성을 매개로 한 아동 발달 장애 진단에 대한 연구가 이루어져 있지만 이전까지는 충분한 샘플을 모으기 힘들다는 한계점이 있었기 때문에 확실한 기준 제시에 대한 부족함이 있다. 그렇기 때문에 주로 정상 군과 하나의 질환 군 분류인 이중 분류의 연구가 진행되어 왔고, 서로 다른 질환에 특이화 된 특성을 알아내기에 많은 어려움이 있었다.¹³⁻¹⁵

아동 발달 장애를 가지고 있는 아이들의 경우, 정상 군의 아이들과 가장 큰 차이점으로 아동 발달 장애가 있는 아이들은 움직임이

제한되거나 오랫동안 집중을 요구하는 상황에서 정상인 아이들에 비해 오랫동안 집중하지 못하고, 움직이고 싶어하기 때문에, 5분 가량의 fMRI촬영 동안에도 뇌에서 다른 형태의 연결성을 떨 것이라고 생각된다.¹⁶ 이 연구에서는 정상 군과 아동 발달 장애 군의 기능적 연결성의 시간에 따른 변화가 큰 차이를 보일 것으로 생각하고, 5분 동안의 연결성의 변화인 Dynamic Functional Connectivity의 평균과 분산을 사용하여 아동 발달 장애 그룹을 분류하는 연구를 진행할 계획이다.

최근에는 다양한 병원 사이의 협동 과제 등으로 많은 데이터를 모아 처리할 수 있는 방법들이 나오고 있어서 부족한 샘플 문제를 해결할 수 있는 상황이 되었다. 하지만 대부분의 빅 데이터가 각기 다른 병원에서 얻은 데이터를 모아놓은 형태이기 때문에 촬영 중 파라미터가 다르게 되고 이런 차이가 분류 모델을 학습할 때 큰 영향을 줄 수 있다.¹⁷ 또한 데이터 획득 과정에서 오류가 생기거나 누락된 데이터가 많이 발생하기 때문에 추후 이것을 생성하거나 보완할 수 있는 방법론이 필요하다.

2. 연구 목적

이 연구에 있어서 가장 중요한 목표는 아동 발달 장애를 겪고 있는 아동을 정확히 분류하는 모델을 만드는 데에 있다. 이 문제를 해결하기 위해서 신뢰도와 응답에 대한 성실성이 부족할 가능성이 있는 설문 응답에 비해 객관적인 데이터가 필요하다. 또 이러한 빅 데이터를 처리함에 있어서

기관마다 파라미터의 차이 보정, 누락 데이터의 생성이 필요한 상황이다. 이 데이터로 학습되어 아동 발달 장애와 정상 군을 분류할 수 있고 추후 새로운 데이터에도 적용할 수 있는 모델을 만드는 것이 이 연구의 가장 큰 목표이다.

3. 연구 범위

아동 발달 장애에 대한 기존에 있었던 진단 방식과 연구보다 더 나은 진단을 하기 위해 2가지 단계의 연구가 필요하다.

첫 번째로, 데이터 처리에 대한 문제를 볼 필요가 있다. 아동 발달 장애 진단에 필요한 rs-fMRI 데이터를 얻을 때, 전 세계의 환자가 각자 다른 병원에서 다른 프로토콜을 사용하여 rs-fMRI 데이터를 얻게 된다. 그렇기 때문에 각 병원마다 이미지의 크기, 촬영 시간, Time Repetition(TR) 등이 다르게 된다. 머신 러닝을 사용하여 진단을 할 때 전 처리가 완료되어 위의 프로토콜과 관계없는 각 관심 영역(ROI) 사이의 기능적 연결성 값을 사용한다. 하지만 각 프로토콜에 차이에 따른 기능적 연결성의 차이가 발생할 수 있고, 차이에 따른 결과의 차이가 어떻게 나타나는지, 보정의 필요가 있는지를 아는 것이 매우 중요하다. 그렇기 때문에 이번 연구에서 중점적으로 다룰 부분으로, 이런 프로토콜의 차이에 기능적 연결성에 차이가 나타나고, 진단에도 차이가 나는지, 프로토콜의 차이와 관계없이 기능적 연결성은 일정하고 진단에도 차이가 없는 지를 연구하는 것이 중요하다고 생각한다. 이 중에서도 촬영 시간의 길이에 대한 연구가 매우 중요하다고 생각되는데 주의력 결핍 및 과잉 행동 장애와 자폐증이 주된

관심진단이기 때문에 촬영 시간을 가능한 한 짧게 사용할수록 원하는 객관적인 데이터를 얻을 수 있다고 생각한다.

두 번째로는, 진단 분류 시 정상 군과 질환 군 만을 분류하지 않고, 질환 군 간의 특성을 알고 분류할 수 있어야 한다. 기존 아동 발달 장애 진단 연구에서는 정상 군과 질환 군 두 가지만을 분류하는 연구들이었다. 하지만 주의력 결핍 및 과잉 행동 장애 같은 경우 Combined type, Inattentive type, Hyperactive type 으로 질환이 세분화 되어있고, 비슷한 증상의 다른 아동 발달 장애들을 구분할 수 있어야 한다. 그렇기 때문에 이중 분류 밖에 할 수 없었던 과거 연구에서 나아가 다중 분류를 할 수 있고, 이중 분류 모델에 비해 정확도가 떨어지지 않은 모델을 만들 필요가 있다.¹⁸

II. 재료 및 방법

1. 데이터 수집

이 연구에 사용할 데이터로는 Child Mind Institute Project(CMI Project)의 뉴욕 3개의 병원에서 수집한 정상 군 집단과 질환 군 집단의 데이터이다. 이 데이터의 총 인원은 1987명으로 55가지의 다른 진단을 가지고 있고, 만 5세부터 18세까지의 연령대가 분포한다. 그 중 이 연구에서는 119명의 주의력 결핍 및 과잉 행동 장애 Combined type과 131명의 주의력 결핍 및 과잉 행동 장애 Inattentive type, 47명의 자폐증, 84명의 정상 군을 선별했다. 이 데이터들은 3개의 다른 기관(Cornell Brain

Imaging Center, Rutgers University, Staten Island)에서 수집한 데이터이기 때문에 각각 다른 파라미터(TR, 스캔 수, 촬영시간 등)를 가지고 있다.

표 1. 기관 별 파라미터 정보

Staten Island (1.5T Siemens Avanto)										
	Slices	%FOV Phase	Resolution (mm)	TR (ms)	TE (ms)	TI (ms)	Flip Angle (deg)	Multi-band Accel	Phase Partial Fourier	Notes
T1	176	100%	1.0x1.0x1.0	2730	1.64	1000	7	Off	Off	
T2	24	87.50%	0.9x0.9x5.0	9000	89	2500	150	Off	Off	
Diffusion Kurtosis Imaging	72	100%	2.0x2.0x2.0	3110	76.2	N/A	90	3	6/8	64 directions, b = 0,1000,2000
fMRI	54	100%	2.5x2.5x2.5	1450	40	N/A	55	3	Off	
MTI	176	100%	1.0x1.0x1.0	30	11	N/A	15	Off	6/8	Acquired with and without MT

Rutgers University Brain Imaging Center (Siemens 3T Tim Trio)										
	Slices	%FOV Phase	Resolution (mm)	TR (ms)	TE (ms)	TI (ms)	Flip Angle (deg)	Multi-band Accel	Phase Partial Fourier	Notes
T1	224	100%	0.8x0.8x0.8	2500	3.15	1060	8	Off	Off	
T2 FLAIR (no longer administered)	22	87.50%	0.9x0.9x5.0	9000	90	2500	150	Off	Off	
T2 SPACE	224	100%	0.8x0.8x0.8	3200	565	Varies	120	N/A	Off	
Diffusion Kurtosis Imaging	72	100%	1.8x1.8x1.8	3320	100.2	N/A	90	3	Off	64 directions, b = 0,1000,2000
fMRI	60	100%	2.4x2.4x2.4	800	30	N/A	31	6	Off	
MTI	176	100%	1.0x1.0x1.0	31	11	N/A	15	Off	6/8	Acquired with and without MT

CitiGroup Cornell Brain Imaging Center (Siemens 3T Prisma)										
	Slices	%FOV Phase	Resolution (mm)	TR (ms)	TE (ms)	TI (ms)	Flip Angle (deg)	Multi-band Accel	Phase Partial Fourier	Notes
T1 - HCP	224	100%	0.8x0.8x0.8	2500	3.15	1060	8	Off	Off	
ABCD T1 w/ vNAV	176	100%	1.0x1.0x1.0	2500	2.88	1060	8	Off	Off	Acquired with and without surface coil intensity correction
T2 SPACE - HCP	224	100%	0.8x0.8x0.8	3200	564	N/A	120	Off	Allowed	
T2 SPACE with vNAV	176	100%	1.0x1.0x1.0	3200	565	N/A	120	Off	Off	Acquired with and without surface coil intensity correction
Diffusion Kurtosis Imaging	81	100%	1.8x1.8x1.8	3320	100.2	N/A	90	3	6/8	64 directions, b = 0,1000,2000
fMRI	60	100%	2.4x2.4x2.4	800	30	N/A	31	6	Off	
MTI	176	100%	1.0x1.0x1.0	32	11	N/A	15	Off	6/8	Acquired with and without MT

CMI 데이터 프로젝트에서 수집한 각기 다른 3개의 병원 뇌 영상 데이터의 파라미터이다. 이 연구에서 사용된 휴지기 fMRI의 파라미터에서 병원마다 TR, 슬라이스 개수 등이 다른 것을 확인할 수 있다.

2. 데이터 전처리

데이터 전 처리의 첫 번째 작업으로는 위의 세 기관에서 수집한 데이터는 fMRI로 촬영된 전 처리가 되어있지 않은 뇌 영상 파일이기 때문에 이 데이터를 분석하기 위해서 전 처리가 필요하다. 전 처리는 SPM12를 사용하여 진행하고, Slicetiming을 하여 시간에 대한 보정, Realignment로 촬영 중간에 움직임과 회전을 보정한다. Co-registration과 Normalization을 통해 개인의 T1영상과 EPI영상을 정합 하고 표준

공간인 MNI space로 변형시켜 준다. 이후 Automated Anatomical Labeling(AAL) map을 사용하여 128가지의 관심 영역(ROI)로 나눠주고 관심 영역 별 혈중 산소 포화도 신호 측정 및 기능적 연결성을 구한다.

두 번째 작업으로 뇌 영상 전처리 소프트웨어인 SPM12로 기본적인 전처리 이후에도 오류를 가지고 있는 영상을 걸러내야 하는 작업을 해야한다. 이번 연구의 뇌 영상 데이터는 질환을 가지고 있는 아동들의 영상이기 때문에 아동들이 5분동안 집중하지 못하고 머리를 움직여 제대로 영상 촬영이 진행되지 않은 경우가 있다. 그렇기 때문에 뇌 EPI 영상을 직접 눈으로 확인하고 그것에서 만들어진 시 계열 데이터와 기능적 연결성을 확인하여 사용할 수 있는 데이터와 오류가 있어서 사용할 수 없는 데이터를 분류하였다.

데이터 전 처리의 마지막 작업으로 관심 영역별 혈중 산소 포화도 신호를 추출하고 기능적 연결성을 구할 때 기관별 파라미터에 대한 차이를 보정해야 한다. 보정해야 하는 파라미터는 다음과 같다.

가. 기관별 Time Repetition(TR) 차이에 따른 보정(Resampling)

나. 기관별 촬영 시간 차이에 대한 보정(5분, 10분)

다. 기관별 스캔 수 차이에 대한 보정(370 스캔, 415 스캔, 750 스캔)

각 파라미터의 차이를 보정하여 동일하게 맞췄을 때의 기능적 연결성과 그렇지 않았을 때를 비교하여 차이가 있는지 보고, 후에 분류 모델에 학습했을 때 차이가 있는 지 확인한다.

3. 기능적 연결성의 변화(Dynamic Functional Connectivity) 계산

이 연구에서는 이미 시 계열 데이터를 1 초의 TR 로 교정했으며 기능적 연결성의 변화를 비교하기 위해 300 초의 일정한 촬영 시간을 사용했다. 촬영 시간이 충분히 길지 않기 때문에 기능적 연결성을 확보하기에 충분히 긴 슬라이스 창 크기를 조정하는 것이 중요하며, 변화의 특성을 보려면 충분한 슬라이스 창의 길이를 확보해야 하기 때문에 적절한 슬라이스 창의 크기가 중요하다. 따라서 슬라이스 창의 크기를 30 초에서 60 초 사이로 설정 한 다음 분류 성능을 비교하여 가장 적합한 창 크기를 30 초로 설정했다. 기능적 연결성의 변화 정도를 얻은 후, 아동 발달 장애가 있는 아동 그룹의 기능적 연결성의 변화가 정상 그룹과 다른지 여부를 결정하기 위해 기능적 연결성 수치 대신 각 슬라이딩 윈도우에 대해 얻은 각 기능적 연결성 변화의 표준 편차를 사용했다.

4. 중요 연결성 추출

90 개의 관심 영역간의 연결성을 계산하는 기능적 연결성에는 4005 개의 너무 많은 연결성이 있으므로 Support Vector Machine(SVM)과 같은 전통적인 분류 모델을 학습하기 전에 그룹 간 차이를 보이는 영역과 연결성 추출을 해야한다.^{12, 19} 연결성 추출을 통해 SVM 학습의 차원을 낮춤으로써 학습 효율성을 높일 수 있으며 분류 후 두 질병 간에 어떤 기능적 연결성이 차이가 있는지 확인할 수 있다.²⁰ 이 연구에서는 SVM 모델의 성능을 중요 연결성 추출 방법인 Multivariate Bayesian

Regression (MVB)²¹ 와 2 그룹 t-test 의 두 가지 방법으로 비교하여 연결성 추출 방법을 선택하기로 결정했다.

5. 분류 모델 구축과 학습

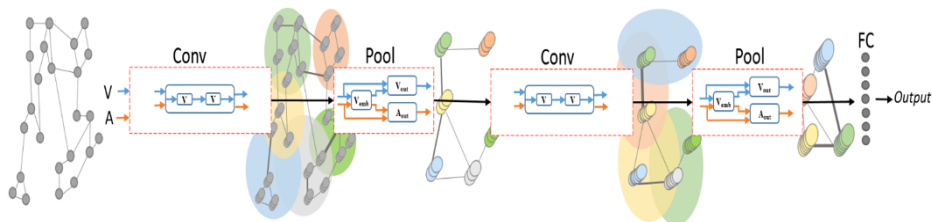
기존 연구에서는 전통적인 방법인 SVM을 주로 사용하여 이중 분류를 하는 연구를 했다. 기존의 연구와 다르게 이번 연구에서는 주의력 결핍 및 과잉 행동 장애의 세부 진단이 많은 점, 정상 군과 주의력 결핍 및 과잉 행동 장애, 자폐증을 서로 분류해야 하는 점이 필요하다고 생각하여 기존 이중 분류인 SVM 뿐 아니라 인공 신경망 모델 중 하나인 Graph Convolution Neural Network(gCNN), Artificial Neural Network(ANN)을 사용하여 다중 분류를 할 수 있는 모델을 결합하여 사용할 예정이다. 네 가지 분류 모델을 결합하여 학습하고 분류함으로써 이중 분류 모델인 SVM의 장점인 높은 성능과 정확도, 인공 신경망의 장점인 다중 분류 두 가지의 장점을 결합할 수 있다.

가. Graph Convolution Neural Network

Graph Convolution Neural Network는 그래프 형태의 데이터를 처리하여 학습하는 구조를 가진다. 최근 많은 연구에서 Graph Convolution Neural Network를 사용한 연구가 진행되었는데 단순한 이미지를 그래프화한 후 Graph Convolution Neural Network에 적용시켜서 학습한 연구로 시작하여, 그래프적 특성을 가지고 있는 웹 데이터나 지문 데이터, 뇌 연결성 데이터 등 그래프 특성을 가지고 있는 많은 구조의 데이터를

적용하여 진행되고 있다.²²⁻²⁴

그림 1. Graph CNN 구조



Graph Convolution Neural Network의 기본 구조로 컨볼루션 레이어와 풀링 레이어로 이루어진 구조를 띄고 있다.²⁵ 기존의 Convolution Neural Network와 가장 큰 차이로 컨볼루션 레이어에서 그래프 데이터의 특성을 많이 활용하는데, 이때 그래프를 이루고 있는 점인 노드와 연결성을 나타내는 엣지에 대한 특성을 반영한 레이어를 만들게 된다.

먼저 그래프 컨볼루션 레이어에서 그래프 특성을 반영하여 전체 노드를 유사한 노드의 특성을 가지고 있는 것 끼리 그룹화를 진행하여 복잡한 입력 값 그래프의 특징이 되는 새로운 그래프를 만들 수 있는 특성을 추출하게 된다. 그래프 컨볼루션 레이어에서 추출된 특성을 바탕으로 바로 다음 연결되는 풀링 레이어에서 데이터 그래프를 단순하게 줄여나가 전체 노드와 엣지를 줄여나가 데이터를 단순화하고 분류 모델의 학습 효율과 성능을 높여준다. 이렇게 단순화 된 그래프를 가지고 분류 학습을 진행하게 되는 구조를 띄고있다.

Graph Convolution Neural Network는 CNN기반의 분류 학습 모델이다. 그 중 점과 점 사이의 연결성을 가지는 그래프를 학습하기 위해 변형된 모델이라고 할 수 있는데, 그래프의 점과 점 간의 연결성을 입력 값으로 하는 머신 러닝 모델의 일종이다.

기존에 사용되고 있는 컨볼루션 레이어와 풀링 레이어는 일반적인 직사각형 그리드에서 사용되기 때문에 그래프 적 특성을 나타내기 쉽지 않다. 그렇기 때문에 Graph Convolution Layer의 가장 큰 목표는 그래프 적 특성을 반영하는 데에 있다.

V 는 노드의 영역을 나타낸 것이고, E 는 가중치가 $W_{i,j}$ 로 정의 된 각 두 노드 영역 간의 연결 여부 이다. 그래프 컨볼루션은 라플라시안 그래프에 의존한다. 라플라시안은 인접한 노드 영역 사이의 신호 변화의 연결성을 특징으로 한다. 정규 라플라스식은 다음과 같이 정의된다.

$$L = I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$$

여기서 D 는 노드 각도의 대각선 행렬이고 I 는 항등행렬이다. 그래프에 정의 된 시 계열 신호인 x 와 필터 g_θ 간의 컨볼루션 연산은 다음과 같이 정의된다.

$$x * g_\theta = UG_\theta U^T x$$

여기서 $G_\theta = \text{diag}(U^T g_\theta)$ 및 $U = (u_0, \dots, u_{N-1})$ 은 라플라시안 고유 벡터의 행렬이다. g_θ 를 다르게 선택하면 다른 그래프 컨볼루션 네트워크를 구성 할 수 있다.²⁶

이 연구에서는 90 개의 뇌 영역별 혈중 산소 포화도 시 계열 데이터를 노드 V 의 입력 값으로 지정하고 노드 간 엣지 값인 $W_{i,j}$ 를 영역간 기능적 연결성 혹은 기능적 연결성 변화의 분산 값으로 지정하고 학습을 진행하였다. 그렇기 때문에 노드 간 시 계열 데이터의 특징과, 엣지 값인 기능적 연결성을 고려하여 모델이 그룹화 되고 레이어를 통과할 수록 단순화 되는 과정을 띄고있다.

이 연구의 Graph CNN 의 구조로는 처음 90 개의 노드와 각각의 연결성인 4005 개의 연결성 값을 입력 값으로 하고 풀링 레이어를 반복할 수록 32 개의 노드, 16 개의 노드로 풀링 되는 구조를 띄고 있다.

나. Support Vector Machine

그룹 분류에는 Support Vector Machine(SVM)을 사용했다.²⁷ SVM은 전통적인 이중 분류 알고리즘으로 다차원에 분포하는 점들 중 경계에 있는 점을 찾아 그 점을 기준으로 경계선을 만들어 내는 방법이다. 이 방법은 학습 시간이 짧고 데이터의 수가 적어도 우수한 성능을 내는 장점이 있는 반면에 두가지의 클래스 만을 비교할 수 있다는 단점을 가지고 있다. SVM에서는 RBF (Radial Base Function) 커널이 분류에 사용되고 RBF 커널은 σ 가 커널의 넓이일 때 다음과 같다.

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)$$

SVM 의 하이퍼 파라미터는 베이지안 최적화 파라미터 방법에 의해 최적화 되었다. 베이지안 최적화 매개 변수는 목적 함수라고 하는 실제 가치 함수를 최소화하는 점을 찾는 프로세스이다. 베이지안 최적화는 내부적으로 목적 함수의 가우시안 프로세스 모델을 유지하고 목적 함수 평가를 사용하여 모델을 훈련시킨다.¹⁹ 목적 함수와 매개 변수를 최적화하기 위해 '예상 개선'방법을 사용했다. 이 방법은 목표 증가의 원인이 되는 값을 무시하고 목표 함수의 예상 개선 정도를 평가한다. 그것은 x_{best} 가 posterior 의 최저 평균의 위치이고, $\mu Q(x_{best})$ 는 그 값일 때, 다음과 같이 정의된다.

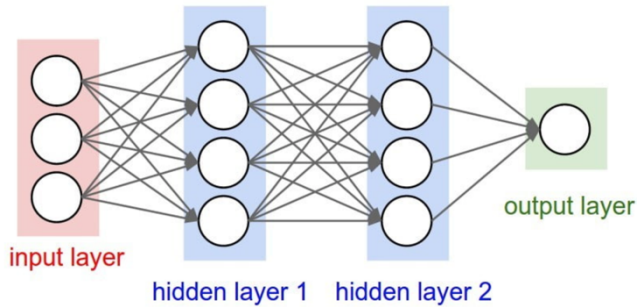
$$EI(x, Q) = E_q[\max(0, \mu Q(x_{best}) - f(x))]$$

이 최적화 방법을 통해 SVM 은 그룹을 적절한 매개 변수로 분류하고 SVM 의 성능을 향상시킨다.

다. Artificial Neural Network

그룹 분류의 세 번째 방법으로 뉴럴 네트워크 방법을 사용하였는데, 뉴럴 네트워크 는 딥 러닝 학습의 기본적인 구조로, 입력 레이어와 출력 레이어 중간에 히든 레이어인 여러 단의 완전 연결 레이어가 자리잡고 있는 구조로 되어있다.²⁸

그림 2. 인공 뉴럴 네트워크 구조



기본적인 인공 뉴럴 네트워크의 구조로, 2개의 완전 연결레이어(히든 레이어)와 입력 레이어, 출력 레이어로 이루어진 딥 뉴럴 네트워크 모델의 구조를 띄고있다.

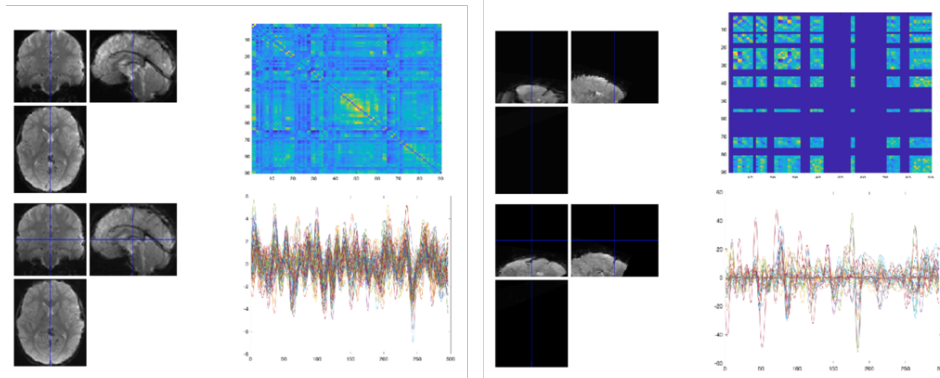
이 연구에서는 각 히든 레이어의 끝의 실행 함수를 sigmoid가 아닌 Relu 함수로 적용하고, 드롭 아웃 레이어를 결합하여 학습 과정에서의 과적합을 방지 하도록 했다.

Sigmoid 함수와 Relu 함수를 비교하면, sigmoid는 값이 음양에 관계없이 결과를 0에서 1사이로 만들어 주는 함수인 반면, Relu함수는 값이 음일 때는 0으로, 양일 때는 값을 그대로 내보내 주는 함수이다.

III. 결과

1. 데이터 전 처리와 오류 검증

그림 3. 데이터 전 처리와 오류 검증



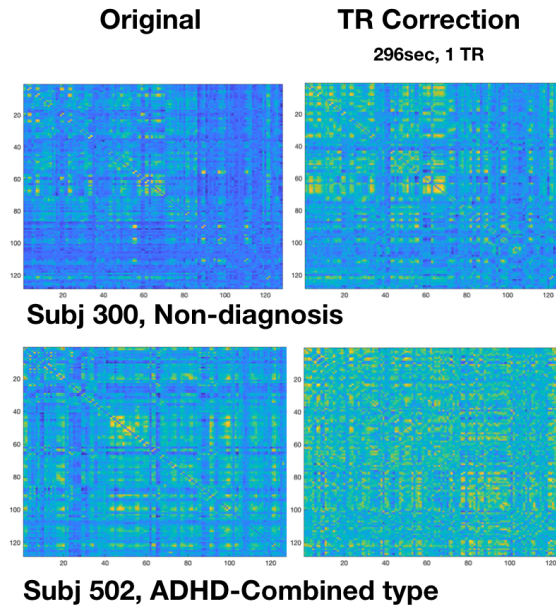
아동 발달 장애의 특성상 움직임이 많은 데이터가 많아서 뇌 영상 이미지와 시 계열 데이터, 기능적 연결성을 구한 후 오류가 있는 데이터를 추출하여 제거하는 작업을 진행함. 왼쪽 그림은 정상적으로 사용할 수 있는 뇌 영상 데이터이고, 오른쪽 그림은 뇌가 움직여서 정상적인 이미지를 얻지 못한 그림이다.

데이터를 사용하는 과정에 있어서 오류가 있는 데이터는 머신 러닝 학습에 치명적인 성능 저하를 일으킬 수 있기 때문에, 데이터 전 처리 작업을 진행하였다. 그림과 같이 오류가 있는 데이터가 다수 발견되었고, 전체의 10%정도의 데이터가 오류를 포함하고 있었다. 그렇기 때문에 이 데이터를 제거하고 다용할 수 있는 데이터만을 사용하여 머신 러닝을 학습하였다. 이 과정 전후의 머신 러닝 분류 성능에서 데이터 전 처리 과정을 진행하기 전 보다 10% 정도의 데이터를 걸러낸 후 성능이

3~4%정도 증가하였다.

2. 기관별 파라미터 보정 결과

그림 4. 파라미터 보정 전 후 기능적 연결성의 차이 비교

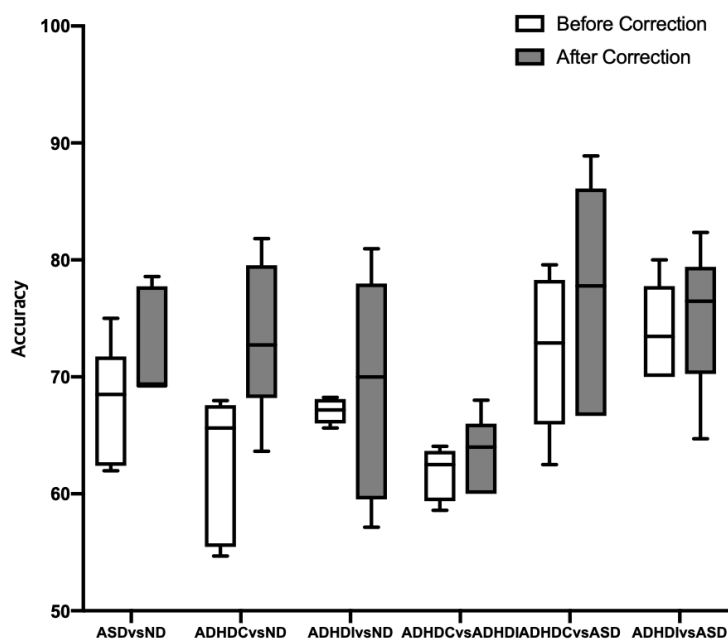


3개의 다른 기관에서 얻은 각자 다른 파라미터를 보정 한 후의 기능적 연결성의 변화를 나타낸 그림이다. 왼쪽 그림은 촬영 시간, TR의 보정을 하지 않고 얻어진 뇌의 128개의 영역의 시 계열 혈중 산소 포화도 신호를 가지고 128개 영역 간 기능적 연결성을 구한 것이고, 오른쪽 그림은 촬영 시간을 296초로, TR을 1로 보정하여 재 생성한 시 계열 데이터를 사용하여 기능적 연결성을 계산한 그림이다. 보정 전 후의 기능적 연결성이 서로 다를 수 있다.

촬영 파라미터 보정으로 인해 데이터의 값이 달라진다는 것을 확인할 수 있고, 어떤 데이터가 아동 발달 장애 진단에 적합한지 아직은 알

수 없기 때문에 더 적합한 데이터를 판단하기 위해서 두 가지 다른 데이터를 사용하여 분류 학습을 진행했다. 이 분류 학습의 성능을 비교함으로써 어떤 데이터가 더 분류를 잘 할 수 있는지 알 수 있기 때문에 두 성능을 비교하여 더 높은 성능을 만들어 내는 데이터를 사용하기로 결정하였다.

그림 5. 파라미터 보정 전 후 분류 학습 성능 비교



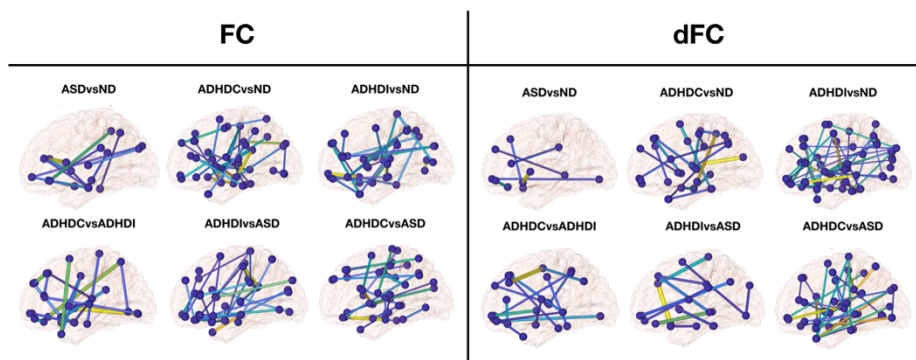
두개의 다른 기능적 연결성 중 어떤 연결성이 그룹 간 차이를 더 잘 나타내고 있는지 확인하기 위해 이 연구에서는 두 가지의 데이터를 Graph CNN에 넣고 분류 학습을 진행하고 분류 성능을 비교하였다. 6가지의 이중 분류 결과를 볼 때 보정을 진행하지 않은 데이터에서는 분류 성능 결과가 70% 전후를 나타내는 데 반해 보정 후의 분류 학습 성능은

최고 87%부터 평균 75%의 성능을 내고 있다.

이 결과를 바탕으로 기관별 파라미터의 차이가 같은 파라미터를 사용하여 얻은 데이터에 비해 아동 발달 장애 진단 성능의 저하를 가져올 수 있다는 결론을 내릴 수 있고, 서로 다른 기관에서 수집한 데이터를 사용하기 위해서는 파라미터의 보정이 반드시 필요하다는 결론을 내릴 수 있다. 또한, 이 결과를 가지고 이번 연구에서도 파라미터가 보정된 데이터를 사용하기로 했다.

3. 중요 연결성 추출 결과

그림 6. 중요 연결성 추출



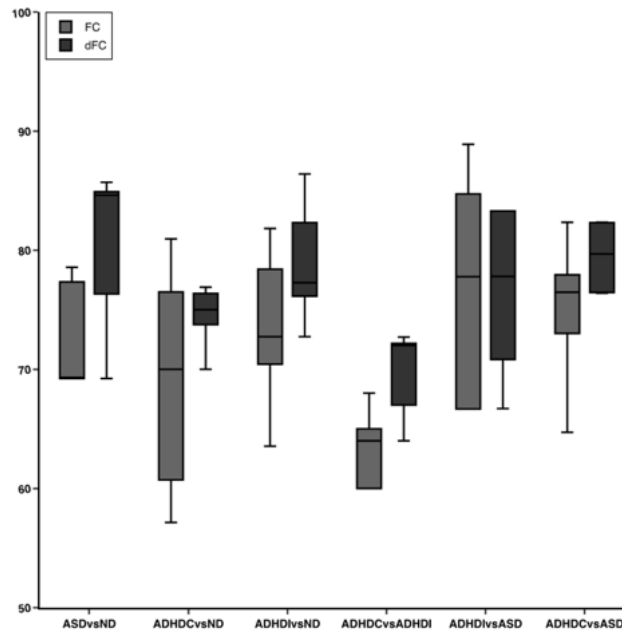
MVB방법으로 기능적 연결성과 기능적 연결성의 변화의 그룹 간 차이를 나타낸 그림이다. 기능적 연결성의 차이와 기능적 연결성의 변화의 그룹간 차이가 다른 영역과 연결성을 띄고 있는 것으로 보아 기능적 연결성의 변화를 가지고 분류를 진행하는 본 연구가 의미를 가진다.

중요 연결성 추출 방법인 MVB 방법으로 기능적 연결성의 중요 연결성 차이를 추출한 결과, 자폐증과 정상 군은 오른쪽 Amygdala 영역과, 왼쪽 Occipital 영역이 큰 차이를 보였다. 주의력 결핍 및 과잉 행동 장애 Combined type과 정상 군은 Precuneus와 Supramarginal 영역이 차이를 보였다. 마지막으로 주의력 결핍 및 과잉 행동 장애 inattentive type과 정상 군에서는 오른쪽 Thalamus 영역에 있는 Putamen과 Caudate 그리고 Temporal 영역이 큰 차이를 보였다.

반면 기능적 연결성의 변화의 중요 연결성의 차이를 MVB로 추출한 결과 자폐증과 정상 군은 왼쪽 Sup Frontal 영역과 중앙 Thalamus 영역이 차이를 보였고, 주의력 결핍 및 과잉 행동 장애 Combined type과 정상 군은 Temporal 영역과 Cuneus 영역이 차이를 보였다. 주의력 결핍 및 과잉 행동 장애 inattentive type과 정상 군에서는 오른쪽 Putamen과 Caudate 영역, Temporal 영역에서 차이를 보였다.

4. 기능적 연결성과 기능적 연결성의 변화로의 분류 성능 비교

그림 7. FC & dFC 분류 성능 비교



기능적 연결성과 기능적 연결성의 변화를 사용한 Graph-CNN 분류 성능 비교이다. 기능적 연결성의 변화를 사용한 분류 성능이 대부분의 비교에서 기능적 연결성을 사용한 분류보다 높게 나왔다.

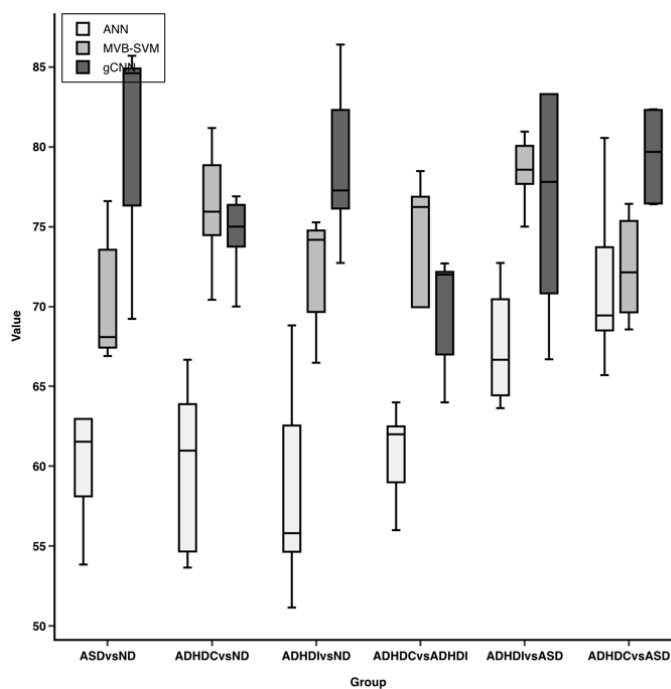
질환 군과 정상 군의 분류 및 질환 군 사이의 분류 과정에서 기능적 연결성의 변화를 사용한 분류가 기능적 연결성을 사용한 분류보다 좋은 성능을 보인다. 기능적 연결성 만을 사용한 분류에서는 최저 60%부터 평균 75%까지의 저조한 분류 성능을 보이는데 반해, 기능적 연결성의 변화를 사용한 분류에서는 최저 78%부터 최고 평균 83.3%까지의 성능을 나타내어 기능적 연결성을 사용한 분류 보다 10%정도 높은 분류 정확도를 보이고 있다.

특히 자폐증과 다른 그룹간의 분류 정확도가 기능적 연결성에 비해 기능적 연결성의 변화를 사용하여 분류한 결과가 더 높은 차이를 보이고

있는데, 이를 바탕으로 자폐증 그룹의 시간에 따른 연결성 변화 정도를 질환 특성으로 생각하고 다른 그룹과의 임상적 비교도 진행해 볼 수 있을 것이다.

5. 분류 모델 별 성능 비교

그림 8. 분류 모델 별 성능 비교



Graph CNN과 MVB-SVM간의 분류 성능을 비교한 그림이다. 주의력 결핍 및 과잉 행동 장애 Combined type을 제외한 다른 모든 그룹 분류 비교 정확도 부분에서 Graph CNN이 전통적인 분류 학습 방식인 SVM보다 더 높은 정확도를 보이고 있다.

위 그림과 같이 인공 신경망 기반인 ANN, Graph CNN과 기존 연구에서 주로 사용되었던 전통적인 방식인 SVM의 분류 모델을 사용하여 정상 군과 아동 발달 장애 군을 분류하고 성능을 비교하였다.

세 가지 분류 학습 모델 중 기본적인 뉴럴 네트워크의 형태를 가진 ANN은 매우 낮은 분류 성능을 보여 6가지의 다른 분류에서 모두 가장 낮은 분류 성능을 보였다. 전통적인 분류 모델인 SVM은 ANN 분류 보다 더 높은 성능을 보였지만 평균 75% 정도의 분류 성능을 보여 두 그룹을 충분히 분류했다고 평가하기 힘든 성능을 보였다. 하지만 Graph CNN 모델은 대부분의 경우에서 세 가지의 분류 성능 중 가장 높은 분류 성능을 보였고, 최고 85%이상의 분류 성능을 보였다.

IV. 고찰

아동 발달 장애는 뇌의 활동과 관련이 있기 때문에 행동의 문제로도 증상이 나타나지만 뇌 기능적 연결성과 연관되어 있다. 그렇기 때문에 아동 발달 장애를 진단하는 데에 있어서 정상 군과 질환 군의 뇌의 기능적 연결성의 차이를 밝혀낸다면, 뇌 활동의 문제가 행동으로 드러나는 간접적인 문제인 행동 문제를 보고 질환을 진단하는 것 보다 더 직접적인 원인을 찾아낼 수 있다. 나아가 문제가 되는 차이에 대한 적절한 자극을 주어서 행동 치료보다 더 효과적인 치료법을 제시할 수 있다고 생각한다. 그렇기 때문에 직접적인 뇌의 연구로 인한 아동 발달 장애 분류가

정확하게 이루어 져야 하고, 그럴 수 있는 모델을 구축하여야 한다고 생각한다.

이 연구는 아동 발달 장애 환자의 특성에 따라서 뇌의 기능적 연결성 뿐 아니라 연결성의 변화로 인한 패턴의 차이를 밝혀냈다는 점에서 기존의 뇌 기능적 연결성 만으로 아동 발달 장애 분류를 진행하던 연구에 비해 아동 발달 장애 행동적 특성을 더 고려한 정보를 사용하여 질환 진단에 반영했다. 실제로 기능적 연결성으로 질환 분류를 하는 것보다 연결성의 변화를 사용하여 분류를 하는 것이 질환 분류가 더 잘된다는 결과를 가지고, 아동 발달 장애 연구에 있어서 휴지기 동안 뇌 영역간의 연결성 자체의 차이보다 변화에서 질환 별 차이가 난다는 것을 밝혀냈다. 기존의 기능적 연결성으로 만들어진 Default Mode Network, Salience Network 등 서브 네트워크와 비교할 수 있는 기능적 연결성의 차이로 인한 서브 네트워크의 생성 등 다양한 방법으로 아동 발달 장애들을 분류할 수 있는 가능성이 있다.

또, 이러한 데이터를 사용하여 머신 러닝 진단을 하여, 보다 자동화된 진단 지표를 만들 수 있다. 아동 발달 장애를 분류하는 데 있어서 전문의가 여러가지 지표를 보고 판단하는 만큼 단기적으로는 전문의의 판단에 도움을 줄 수 있는 지표를 만들어 냄으로써 더 정확한 진단을 도와줄 수 있고, 장기적으로는 머신 러닝의 분류가 더 높은 성능으로 아동 발달 장애를 분류할 수 있게 되면 곧바로 진단으로 활용될 수 있다.

이 연구는 여러 기관에서 다른 파라미터를 사용하여 얻어진 뇌 영상 빅 데이터를 전처리 하여 TR, 촬영 시간 등을 보정하여 사용하였다는

점에서 여러 기관에서 모은 빅 데이터 사용과 전처리 과정에 도움을 줄 수 있다고 생각한다. 한 기관에서 모으고 학습하지 않고 여러 기관에서 협력을 통한 데이터 사용과 추가되는 데이터를 분류하기 위해서 반드시 필요한 과정을 거쳤고, 다른 데이터들을 하나로 모으는 방법을 제시하고 있다.

이 연구에서 아동 발달 장애 진단에 대한 연구를 진행하였지만, 더 시도해볼 연구가 많이 남아있다고 생각한다. 질환이 있는 아동들을 대상으로 하는 연구인 만큼, 아동들이 짧은 시간 동안 집중할 수 있게 하는 것이 매우 중요하다. 그렇기 때문에 기존에 5분동안 촬영된 fMRI영상을 가지고 그룹 비교를 하던 것에서 나아가 더 짧은 시간의 시 계열 데이터를 가지고 질환 진단 분류 학습을 진행하여 아동들에게서 얻을 수 있는 신뢰할 수 있는 데이터의 길이와 기능적 연결성을 알아내는 것이 가장 중요하게 진행해야 할 부분이라고 생각한다.

그 다음으로 진행해야 할 중요한 주제는 다 그룹간 비교 분류이다. 이중 분류를 주로 하는 전통적인 방법인 SVM에서 벗어나 Graph CNN 모델을 구축한 만큼 다중 분류 학습을 진행하여 정상 군과 주의력 결핍 및 과잉 행동 장애의 2가지 하위 질환 군 간 3그룹 분류 학습을 진행해야 하고, 더 나아가서 자폐증을 포함하여 4그룹 간 동시 분류 학습을 진행할 필요가 있다.

아동 발달 장애 연구에는 두 가지 한계가 있다. 첫째, 정확한 진단을 내릴 수 있는 지표가 없다는 점이다. 현재 아동 발달 장애 진단에 있어서 진단을 하는 방법으로, 임상에서 전문의가 DSM-5 외에도 설문지

기본 인덱스 설정 및 진단을 위해 바이오 마커를 사용하여 진단을 내리고 있지만, 정확하게 아동 발달 장애를 수치적으로나 기준점을 제시할 수 있는 지표인 'Golden Standard'가 존재하지 않는 상황이다. 현재 진단되고 있는 아동 발달 장애들도 높은 수준의 지표와 전문의의 진단으로 이루어지고 있기는 하지만 다양한 방법을 제시하면서 더욱 정확한 진단을 하기 위해 나아가고 있는 과정이기 때문에 현재 데이터가 가지고 있는 질환 진단을 절대적인 기준으로 생각하기 힘들다는 문제가 있다.

두 번째로, 비슷한 증상을 가진 질병이 많고 아동 발달 장애 수준의 차이로 인해 어떤 기준부터 아동 발달 장애로 진단할 수 있는 기준이 모호하다. 정상 군 범주에 속하지만 고 위험 군, 질환 군 범주에 속하지만 약한 증상의 질환 정도를 가지고 있는 군 등 다양한 등급의 질환 정도가 있기 때문에 정상 군과 질환 군 두가지로 나누기 힘든 상태이다. 그렇기 때문에 지금 사용되고 있는 진단이 정확하게 두 그룹으로 확실하게 나뉘어져 있지만 고위험 군과 약한 증상의 질환 군 사이의 뇌 기능적 연결성의 차이가 크게 나타나지 않는 경우도 있고, 실제로 기능적 연결성으로의 판단은 기존 설문 지표로 진단된 그룹과의 차이가 생길 수 있다.

아동 발달 장애 진단 연구에는 이러한 한계점들이 있기 때문에 정확한 분류 결과는 만들어 내는 것이 쉽지 않은 상황이다. 하지만 이 연구와 같이 휴지기 기능적 연결성 등 기존 진단 방법 외의 지표를 활용한 연구가 지속되고 다양한 각도에서의 분류가 진행된다면 더 정확한 아동 발달 장애를 분류해 낼 수 있을 것 이라고 생각한다.

V. 결론

이번 연구의 가장 큰 목표는 아동 발달 장애라는 행동적 점수, 설문 지표로 정확한 진단이 어려운 질환을 더 정확하게 진단하는 데에 있어서 객관적인 바이오 마커인 기능적 연결성을 사용하여 머신러닝 자동 분류 학습을 하는 데에 있었다. 질환 군 사이에 비슷한 행동적 특성을 나타내기도 하고, 명확하게 경계선이 존재하지 않는 분류를 해내기 위해 뇌 발달 장애의 근본적인 문제인 뇌 활동의 차이를 정상 군과 분류하고, 차이가 있는 영역간의 활동을 추출함으로써 질환 군의 특성을 확실히 나타내기 위한 연구를 진행하였다.

이 과정 중 데이터 수집에 있어서 전 세계적 추세인 여러 기관의 협력으로 얻어진 데이터를 사용하였고, 그에 따라 발생하는 문제점인 기관별 파라미터 차이와 그것의 보정에 대한 부분에 여러가지 시도를 하였고, 성능을 높이기 위한 방법을 찾아내었다. 이 과정에서 기관 별 다른 파라미터를 하나로 보정하는 작업이 매우 중요하고 분류 학습의 성능을 높이는 데에 큰 도움이 된다는 것을 발견하였다.

아동 발달 장애를 진단하는 데에 있어서 객관적인 바이오 마커인 기능적 연결성을 사용한 연구를 진행하였지만 아동 발달 장애의 특성인 많은 활동과 변화에 대한 특성을 데이터에 접목시키기 위해서 이 연구에서는 fMRI 촬영 동안 짧은 기간 별 기능적 연결성이 변화하는 것에 집중하여 30 초 동안의 짧은 시간 동안의 기능적 연결성을 계산하여 촬영

시간 동안 기능적 연결성 변화 크기에 대한 계산을 하고 그것을 사용함으로써 아동 발달 질환의 특성을 고려하여 분류 학습을 진행하였다.

기존 분류 학습 연구에서는 전통적인 분류 학습 모델인 서포트 벡터 머신 이나 디시전 트리 등을 사용하여 분류 학습을 진행하였다. 하지만 이번 연구에서는 뇌의 영역간 연결성의 정도를 계산하여 만들어진 데이터를 사용했기 때문에, 데이터의 특성을 고려하여 데이터에 더 잘 맞는 분류 모델인 Graph Convolution Neural Network 를 사용하여 분류 학습을 진행하였다. 이 모델을 사용하여 분류 학습을 진행하고 기존 연구에서 주로 사용되던 전통적 분류 학습 모델 중 하나인 서포트 벡터 머신 과의 성능 비교를 통해 데이터의 특성을 활용한 Graph Convolution Neural Network 모델이 서포트 벡터 머신 보다 더 높은 분류 성능을 나타내는 것을 확인하였다.

참고문헌

1. Du Y, Fu Z, Calhoun VD. Classification and prediction of brain disorders using functional connectivity: Promising but challenging. *Frontiers in neuroscience*. 2018;12.
2. Murias M, Swanson JM, Srinivasan R. Functional connectivity of frontal cortex in healthy and ADHD children reflected in EEG coherence. *Cerebral Cortex*. 2006;17(8):1788-1799.
3. Tian L, Jiang T, Wang Y, Zang Y, He Y, Liang M, et al. Altered resting-state functional connectivity patterns of anterior cingulate cortex in adolescents with attention deficit hyperactivity disorder. *Neuroscience letters*. 2006;400(1-2):39-43.
4. Tomasi D, Volkow ND. Functional connectivity of substantia nigra and ventral tegmental area: maturation during adolescence and effects of ADHD. *Cerebral cortex*. 2012;24(4):935-944.
5. Assaf M, Jagannathan K, Calhoun VD, Miller L, Stevens MC, Sahl R, et al. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage*. 2010;53(1):247-256.
6. Tian L, Jiang T, Liang M, Zang Y, He Y, Sui M, et al. Enhanced resting-state brain activities in ADHD patients: a fMRI study. *Brain and Development*. 2008;30(5):342-348.
7. Czerniak SM, Sikoglu EM, King JA, Kennedy DN, Mick E, Frazier J, et al. Areas of the brain modulated by single-dose methylphenidate treatment in youth with ADHD during task-based fMRI: a systematic review. *Harvard review of psychiatry*. 2013;21(3):151.
8. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine*. 2017;23(1):28.
9. Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, et al. Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage*. 2018;169:431-442.
10. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*. 2018;17:16-23.
11. Dey S, Rao AR, Shah M. Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects. *Frontiers in neural circuits*. 2014;8:64.
12. Ng AY. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. Paper presented at: Proceedings of the twenty-first international conference on Machine learning, 2004.
13. Anderson A, Douglas PK, Kerr WT, Haynes VS, Yuille AL, Xie J, et al. Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. *NeuroImage*. 2014;102:207-219.

14. Hart H, Chantiluke K, Cubillo AI, Smith AB, Simmons A, Brammer MJ, et al. Pattern classification of response inhibition in ADHD: toward the development of neurobiological markers for ADHD. *Human Brain Mapping*. 2014;35(7):3083-3094.
15. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: Classification and criteria changes. *World Psychiatry*. 2013;12(2):92-98.
16. Preti MG, Bolton TA, Van De Ville D. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*. 2017;160:41-54.
17. Yamashita A, Yahata N, Itahashi T, Lisi G, Yamada T, Ichikawa N, et al. Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLoS biology*. 2019;17(4):e3000042.
18. Dai D, Wang J, Hua J, He H. Classification of ADHD children through multimodal magnetic resonance imaging. *Frontiers in systems neuroscience*. 2012;6:63.
19. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. Paper presented at: Proceedings of the 25th international conference on Machine learning, 2008.
20. Chen CP, Keown CL, Jahedi A, Nair A, Pflieger ME, Bailey BA, et al. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *NeuroImage: Clinical*. 2015;8:238-245.
21. Lee D, Yun S, Jang C, Park H-J. Multivariate Bayesian decoding of single-trial event-related fMRI responses for memory retrieval of voluntary actions. *PloS one*. 2017;12(8):e0182657.
22. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*. 2016;30(8):595-608.
23. Ying R, He R, Chen K, Eksombatchai P, Hamilton WL, Leskovec J. Graph convolutional neural networks for web-scale recommender systems. Paper presented at: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018.
24. Zhou J, Cui G, Zhang Z, Yang C, Liu Z, Wang L, et al. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*. 2018.
25. Zhang Y, Bellec P. Functional Decoding using Convolutional Networks on Brain Graphs.
26. Such FP, Sah S, Dominguez MA, Pillai S, Zhang C, Michael A, et al. Robust spatial filtering with graph convolutional neural networks. *IEEE Journal of Selected Topics in Signal Processing*. 2017;11(6):884-896.
27. Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995;20(3):273-297.
28. Deshpande G, Wang P, Rangaprakash D, Wilamowski B. Fully connected

cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE transactions on cybernetics*. 2015;45(12):2668-2679.

ABSTRACT

Optimization of Deep Neural Network for Heterogeneous Big Cohort Data in the Classification of Neurodevelopment Brain Disease

Hyunseok Bahng

*Department of Medical Science
The Graduate School, Yonsei University*

(Directed by Professor Hae-Jeong Park)

Diagnosis of child developmental disorders, such as attention deficit and hyperactivity disorder (ADHD) or autism (ASD) based on previous survey responses, has raised questions about reliability and responsiveness. Therefore, Diagnosis studies have continued with biomarkers such as functional connectivity. However, in the case of disease diagnosis through machine learning, in the past, there was a limitation of lack of a sample for model training, and there is a lack of research for subtype of one diagnosis or classification between diagnosis. These multiple classifications must be studied because creating criteria for diagnosing multiple groups at once, rather than the normal and disease groups, is an important way to identify the characteristics of a disease. To overcome the limitations of previous research, two studies were conducted, each with preprocessing corrections for data obtained with different parameters in the hospital, and a good machine learning classification model. In addition, the change in resting functional connectivity is calculated by considering the behavioral characteristics of the child's developmental disability, and the model training and classification is conducted to compare the performance of the model with functional connectivity. In this process, we also compared the performance of the traditional classification model such as support vector machine and using a deep artificial neural network with the characteristics of the graph structure connected by functional connectivity between regions.

Key Words: neurodevelopment disease, graph CNN, SVM