

뇌파 신호 기반 BCI 연구에서 데이터 연속성의 영향

김윤상¹, 한주혁¹, 김웅식^{2*}

¹건양대학교 의료공학과, ²건양대학교 의료인공지능학과

Impact of Data Continuity in EEG Signal-based BCI Research

Youn-Sang Kim¹, Ju-Hyuck Han¹, Woong-Sik Kim^{2*}

¹Department of Medical Engineering, Konyang University

²Department of Medical Artificial Intelligence, Konyang University

요약 본 연구는 시계열 데이터의 연속성과 인공지능 모델의 분류 성능에 대한 비교 실험을 수행하였다. EEG 신호를 이용한 BCI 연구에서는 데이터 연속성이 감소할수록 행동과 사고 분류의 성능이 향상되었다. 특히, LSTM은 연속성이 낮은 데이터에서 0.8728이라는 높은 성능을 달성하였고, 연속성을 고려하지 않은 경우 DNN이 0.9178의 성능을 보였다. 연속성을 고려하지 않은 데이터가 더 우수한 성능을 보일 수 있음을 시사하였다. 또한, 연속성을 고려하지 않은 데이터는 작업 분류에서도 더 높은 성능을 보였다. 이러한 결과는 뇌파 신호를 기반으로 한 BCI 연구에서는 데이터 연속성을 고려하기보다는 셔플링을 통해 다양한 데이터 특성을 보여줌으로써 우수한 성능을 발휘할 수 있음을 시사한다.

• 주제어 : EEG 신호, BCI 연구, 시계열 데이터, 분류 성능 비교, 데이터 연속성

Abstract This study conducted a comparative experiment on the continuity of time series data and the classification performance of artificial intelligence models. In BCI research using EEG signals, the performance of behavior and thought classification improved as the continuity of the data decreased. In particular, LSTM achieved a high performance of 0.8728 on data with low continuity, and DNN showed a performance of 0.9178 when continuity was not considered. This suggests that data without continuity may perform better. Additionally, data without continuity showed better performance in task classification. These results suggest that BCI research based on EEG signals can perform better by showing various data characteristics through shuffling rather than considering data continuity.

• Key Words : EEG Signal, BCI research, Time series data, Classification Performance, Data continuity

Received 19 December 2023, Revised 27 March 2024, Accepted 30 March 2024

* Corresponding Author Woong Sik Kim, Department of Medical Artificial Intelligence Konyang University, 158, Gwanjeodong-ro, Seo-gu, Daejeon, Korea. E-mail: wskim@konyang.ac.kr

I. 서론

뇌파는 뇌의 뉴런에서 발생하는 미약한 전기 신호로, 수 만 개의 뉴런에 전파되면서 측정 가능한 정도의 전기 신호를 생성한다[1]. 뇌파는 전위가 연속적으로 변하는 시계열 데이터이며, 주로 다섯 가지 주파수 범위로 나뉜다. 델타(δ)파는 주파수 0.2~4Hz로 깊은 수면이나 명상 상태, 의식 불명 상태에서 나타난다. 세타(θ)파는 주파수 4~8Hz로 창의적인 생각이나 정서 안정, 잠에 빠져드는 과정에서 관찰된다. 알파(α)파는 주파수 8~13Hz로 편안하고 안정된 상태에서 발생하며, 눈을 감거나 편안한 상태에서 알파파가 나타난다.

뮤(μ)파는 알파에 속하며, 7.5~12.5Hz(주로 9~11Hz)에서 신체의 물리적인 움직임과 관련이 있다. 베타(β)파는 주파수 13~30Hz로 일상적인 인지 및 사고 활동에서 발생하며, 복잡한 계산이나 많은 생각을 할 때 주로 나타난다. 감마(γ)파는 주파수 30Hz 이상으로 극도의 긴장이나 흥분 상태, 고집중 상태에서 관찰된다[2].

뇌-컴퓨터 인터페이스(BCI, Brain-Computer Interface)는 뇌에서 발생하는 뇌파를 측정하여 생각만으로 컴퓨터나 기계를 제어하는 기술로, 다양한 분야에서 지속적으로 연구되고 있다. 이러한 BCI 기술은 연속적인 뇌파 데이터를 사용하여 사용자의 의도를 더욱 정확하게 인식할 수 있다[3].

Lo Giudice, et al.[4] 는 운동 장애를 갖는 사람을 위한 안구 운동 장치제어를 목적으로 수행되었다. 이 연구는 EEG 신호를 기반으로 안구의 자발적 깜박임과 비자발적 깜박임을 자동 분류하기 위한 1D-CNN 기반의 인공지능 분류모델을 제안했으며, 이 모델은 97.92%의 성능을 보였다.

Cheng, S., et al.[5] 은 뇌파와 안구 운동 데이터를 효과적으로 융합하는 방법을 제안했다. EEG와 안구 데이터를 융합한 모터 이미지 분류 정확도가 순수 EEG 데이터 또는 순수 안구 운동 데이터보다 더 높은 것으로 나타났다.

Lee, H.S., et al.[6] 는 uHD EEG인 g.Pangolin을 사용하여 한 손의 개별 손가락 움직임을 디코딩하여 성능을 탐색하는 연구를 진행하였다. 5개의 단일 손가락 움직임을 디코딩 하였을 때 평균 정확도는 각각 70.7%, 65.3%, 63.7%, 63.7%, 60.6%의 정확도를 보였다.

Shin, J., et al.[7] 은 암산, 운동 행동 상상, 그리고 휴식 상태에서 파생된 뇌 활성화 패턴의 정보 전송 속

도를 높이기 위해 hBCI(hybrid-BCI)의 사용을 제안하였다. 데이터를 필터 बैं크 기반의 정규화된 공동 공간 패턴을 EEG 데이터에 적용하여 전처리를 한후 EEG-BCI, NIRS-BCI, hBCI 에 대한 평균 분류 정확도를 sLDA(Shrinkage Linear Discriminant Analysis) 를 사용하여 분류하였다.

Chai, R., et al.[8] 은 ICA(Independent Component Analysis) 기반의 BSS(Blind Source Separation) 기법을 활용하여 암산과 관련된 작업을 분류하는 연구를 수행했다. 데이터를 PSD(Power Spectrum Density) 기반으로 2채널 뇌파 데이터 특징 추출 후 베이지안 신경망으로 분류하였다. 6개의 EEG 채널(C3, C4, P3, P4, O1, O2)에서 2개의 채널을 융합하여 각각 정확도를 분류하였으며 O1&C4에서 가장 높은 정확도 76.4%를 보였다.

Darmakusuma, R., et al.[9] 은 휴식 상태와 움직임 발생 전 상태를 구별하여 움직임을 예측하였다. 7~30Hz의 EEG 신호를 사용하여 움직임 발생 전 상태와 휴식 상태를 구분하였다. Mu와 Beta의 리듬을 평균 TPR(True Positive Rate) 0.64 ± 0.11 와 평균 FPR(False Positive Rate) 0.17 ± 0.08 을 통해 확인하였다.

Huang, C., et al.[10] 은 EEG 신호 분류의 정확성 향상을 위한 Multi-Scale CNN 모델을 제안한다. 이 연구에서는 BCI Competition IV Dataset 2b를 사용했으며, Layer 계산에 따라 발생하는 CNN Feature map의 소실을 문제로 정의하여 Skip-Connection 형태의 Layer 연결을 적용했다. 이 방법의 검증은 ANN(Artificial Neural Network), SVM(Support Vector Machine), SAE(Stacked Auto-Encoder)와 비교 실험을 통해 수행되었으며, Multi-Scale CNN이 평균 73.9%로 높은 성능을 보였다.

Cho, J. H., et al.[11] 은 EEG 신호로부터 근육 활동 패턴을 추정하여 분류 정확도를 향상시키는 접근 방식을 제안했다. 제안 모델과 CSP(Common Spatial Pattern)와 LDA(Linear Discriminant Analysis)를 포함하는 모델 I과 정규화된 필터 बैं크와 CSP를 포함하는 모델 II를 비교하여 실험을 진행하였으며 제안된 모델의 정확도가 63.89%로 나왔으며 모델 I과 모델 II의 평균 정확도도 각각 24.01%, 21.59% 높은 것으로 나타났다.

뇌파 연구에 있어 기존 연구에서 어떠한 뇌파를 활용하여 연구를 진행했는지 비교 분석을 통해 우리의 연구에서 사용할 5가지 행동을 선정하였다. 행동 선정 이후 기존 연구에서 우리의 연구에 적합한 1D-CNN,

MS-CNN, ANN(DNN)을 선정하였고 뇌파가 시계열 데이터인 점과 연속성을 고려한 데이터에 대한 분류 작업을 위해 LSTM, GRU 모델을 선정하여 총 5가지 모델을 선정하여 연구를 진행하였다.

기존 연구에서는 뇌파 데이터의 부족으로 Public Data를 활용하기 어려웠고, 연구자들은 Manual Annotation을 통해 데이터를 획득하였다. 이로 인해 데이터의 특성과 전처리 방법, 연속성이 표준화되지 않아 모델 성능 비교의 어려움이 있다. 우리 연구는 이 문제를 해결하기 위해 5가지 행동에서 얻어진 뇌파 데이터를 통일된 전처리와 데이터 연속성으로 구성하여 BCI 모델 성능을 비교 실험을 진행 하였다. 모델의 하이퍼 파라미터와 복잡도를 최대한 비슷한 환경으로 구성하였다. 이로써 모델 간 공정한 비교와 연구 결과의 일반화 가능성이 높아졌다. 연구는 EEG 데이터인 시계열 데이터의 연속성에 따른 데이터 구성을 3가지 방법으로 나누어 연구를 진행했으며 5가지의 인공지능 모델에서 비교 분석을 하였다. 우리의 연구를 통해 시계열 데이터의 연속성이 EEG 뇌파 분류 성능에 미치는 영향을 확인하고 다양한 데이터셋에 따른 최적의 분류 모델을 식별하고자 노력하였다. 연구의 결과는 다른 BCI 연구에 유용한 정보를 제공할 것으로 기대한다.

II. 연구 방법

2.1 Datasets

본 연구에서는 EEG 신호를 기반으로 수행되는 BCI 연구들의 비교 분석을 위한 실험을 수행했다. 실제 인간 대상의 비침습식 뇌파 측정기(Quick-20r from CGX A cognionics company)를 통해 뇌파 데이터를 수집하였다. 모든 피실험자에게 실험의 목적과 과정을 설명하였으며, 동의서에 서명하였다. 건양대학교 생명윤리 위원회(IRB)의 승인 하에 연구를 진행 하였다.(KYU 2022-12-004-001). 본 데이터의 수집 전극은 10-20 시스템을 기준 측정점으로 정하여 전극을 배치하고 뇌파를 수집하였다. 실험에 참여한 피험자는 25명으로 구성 되었다. 19 ~ 60세 사이의 건강한 성인 남성 11명, 여성 14명으로 구성 되었으며 평균 연령은 28세이다. 분류를 수행하기 위해 5가지 행동을 취하여(눈 뜬 상태, 눈 감은 상태, 왼손을 흔드는 상태, 오른손을 흔드는 상태, 암산 상태) 데이터를 획득 및 데이터 세트를 구성한다. 행동 상태 마다 30초씩 뇌파 수집을 수행해

한 명의 피험자에게 총 2분 30초의 신호가 도출된다. 오류 방지를 위해 데이터의 시작과 끝 각 1.5초를 제외한 27초 분량의 데이터를 사용했다. 각 행동 별로 총 675초의 행동 뇌파를 사용하여 연구를 진행하였다. 사용된 뇌파는 모든 행동을 포함하여 3,375초 간의 뇌파 데이터가 수집되어 활용되었다.

2.2 Preprocessing

본 연구에서는 데이터 수집을 통해 획득한 데이터 세트를 전처리한다. Sampling Rate는 500Hz이며 EEG Channel은 21Ch이다. 획득된 뇌파 데이터는 한 행동당 30초간 0.002초 간격으로 측정이 되었다. 데이터는 Notch-Filter를 걸쳐 60Hz의 전원 노이즈를 제거하였다. 이후 Band-Pass Filter를 통해 데이터를 4Hz ~ 59Hz 사이로 필터링 하였다.

2.3 Experiment

본 연구에서는 데이터 구성에 따라 시계열 데이터의 연속성을 고려한 데이터를 평가를 위해 실험 데이터 균을 3가지로 나누어 구성한다. Table 1은 본 연구에서 실험을 위해 구성한 데이터이며, 데이터는 연속 범위와 랜덤 셔플을 기준으로 6개로 분류된다.

Table 1. Configuration of Datasets

	Merge	RS [†] ₁	SS [†] ₂ _500	SS _250	SS _125	SS _25
*CR (sec)	150	0	1	0.5	0.25	0.05
**RS (T/F)	F	T	T	T	T	T

*CR : Continuous Range, **RS : Random Shuffling

RS[†]₁ : Random Sample, SS[†]₂ : Slicing Sample.

연구에서는 Merge, Random Sample, Slicing Sample 세 가지 유형의 데이터셋을 사용하여 5가지 분류 모델(DNN, LSTM, 1D-CNN, GRU, MS-CNN)을 실험을 진행 하였다. Merge 데이터는 연속성을 유지한 형태, Random Sample은 과적합과 데이터 편향 방지를 위해 무작위로 샘플된 형태, Slicing Sample은 연속성을 고려하여 일정 부분을 잘라내고 섞은 형태이다. 실험은 초 단위에 따라 4가지(1s, 0.5s, 0.25s, 0.05s)로 분류되었다. Merge와 Random Sample 데이터셋에서는 DNN, LSTM,

1D-CNN을 활용하였고, 연속성을 고려한 Slicing Sample에서는 GRU와 MS-CNN을 추가로 실험하였다. GRU의 경우 모델의 복잡도에 따른 성능 비교를 하기 위함과 동시에 추후 실시간성을 고려한 연구를 진행할 때 뇌파 측정 기기와 실시간 학습 속도를 고려하기 위해 선정하였으며, MS-CNN의 경우에도 복잡도에 따른 성능 비교와 더불어 다양한 스케일의 정보에서의 비교 분석을 진행하기 위해 선정하여 진행하였다. 모든 모델은 동일한 환경에서 비교 분류 실험을 진행하여 각 모델의 데이터 유형에 따른 성능을 비교하고, GRU와 MS-CNN이 연속성을 고려한 데이터에서 어떠한 강점을 가지는지 확인하였다.

연구는 최종적인 목표인 실시간 BCI 구현을 목표로 하여 연구를 진행하였다. 실시간성을 보장하기 위해서는 분류의 속도에서 빠른 분류 속도가 보장이 되어야 하기 때문에 모델 구조를 비교적 간단하게 구성하여 연구를 진행하였다. 모델들은 공통적으로 각 모델의 구조를 거친 후 Flatten 작업을 거쳐 Dense Layer에서 다중 분류 작업을 진행한다.

Table 2. Test Model Configuration by Datasets

	Merge	RS	SS _500	SS _250	SS _125	SS _25
DNN† ₁	295,734	295,734	288,755	289,359	294,254	292,397
LSTM† ₂	292,357	292,357	292,437	291,305	291,920	293,270
1D-CNN† ₃	291,549	291,549	293,817	292,407	291,957	291,957
GRU† ₄	-	-	293,561	292,901	292,521	292,245
MS-CNN† ₅	-	-	299,175	299,385	290,845	294,205

DNN†₁ : Deep Neural Network, LSTM†₂ : Long Short Term Memory, 1D-CNN : 1-Dimensional Convolutional Neural Network, GRU†₄ : Gated Recurrent Unit, MS-CNN†₅ : Multi-Scale CNN

Table 2는 데이터 세트에 대한 실험에서 사용된 분류 모델의 복잡도를 보여준다. DL 분류기들은 약 300,000의 유사한 복잡도를 가지며, 입력 데이터에 따라 계층 구성이 약간씩 다르다. 이 차이는 미세한 복잡도 차이를 유발하며, 데이터 세트의 연속성을 손상시키지 않고 목적에 맞게 모델에 적용된다. LSTM과 GRU 같은 RNN 계열 분류기는 순차적인 데이터에서 Feature의 상태 벡터를 저장하여 다음 벡터에 영향을 끼치는 강점을 가지고 있다. Merge 및 RS 데이터셋에

서는 GRU와 Multi-Scale CNN에 대한 실험이 수행되지 않았고, Slicing Sample 데이터셋에서는 GRU와 Multi-Scale CNN이 연속성을 고려한 데이터 분류에서의 강점을 확인하기 위해 실험되었다.

III. 결과

본 연구에서는 데이터 구성에 따른 연속된 Feature 가 분류 모델 별로 어떠한 성능을 보이는지에 대한 실험을 수행했다. 이를 위하여 실험 데이터 군을 3가지로 나누어 구성하였으며, BCI 분류 문제에 대한 대표적인 DL 분류기를 선정하여 비교했다. 모델을 비교할 때에 같은 조건에서 비교하기 위해 하이퍼 파라미터를 동일하게 설정하여 연구를 진행하였다.

Table 3. Hyper Parameter Setting

Hyper parameter	Common Parameter Setting
Optimizer	Adam
Learning Rate	0.0001
Loss Function	Categorical Crossentropy
Activation Function	Softmax

Table 3과 같이 하이퍼 파라미터를 설정하여 연구를 진행하였다. 또한 학습을 진행하며 일어날 수 있는 과적합을 막기 위해 EarlyStopping을 통해 Validation Loss를 모니터링 하였으며 Epoch가 증가할수록 학습률을 감소하는 지수 함수를 사용하였다.

Table 4. Deep Learning classifier Performance on Merge Dataset

Matrix	Model	Eyes Open	Eyes Closed	Waving left hand	Waving right hand	Mental arithmetic
Precision	DNN	0.9258	0.9278	0.9101	0.9091	0.9122
	LSTM	0.9288	0.9258	0.9026	0.9028	0.8767
	1D-CNN	0.9263	0.9081	0.9071	0.9029	0.9094
Recall	DNN	0.9412	0.9303	0.9112	0.8984	0.9040
	LSTM	0.9169	0.9109	0.9028	0.8911	0.9132
	1D-CNN	0.9266	0.9263	0.9118	0.8966	0.8925
F1-Score	DNN	0.9334	0.9291	0.9107	0.9037	0.9081
	LSTM	0.9228	0.9183	0.9027	0.8969	0.8946
	1D-CNN	0.9264	0.9172	0.9094	0.8997	0.9009

Table 4는 Merge Dataset에 대한 DL 분류기 성능을 보여준다. Precision에서 DNN이 주로 뛰어나며, LSTM은 Eyes open 상태에서 우수한 결과를 보였다. Recall에서는 DNN이 Eyes open 및 closed, Waving right hand에서 우수하며, Waving left hand에서는 1D-CNN이, Mental arithmetic에서는 LSTM이 우수하다. F1-Score에서는 DNN이 모든 분류에서 우수한 성능을 보인다.

Table 5. Deep Learning classifier Performance on RS Dataset

Matrix	Model	Eyes Open	Eyes Closed	Waving left hand	Waving right hand	Mental arithmetic
Precision	DNN	0.9361	0.9311	0.9156	0.8965	0.9102
	LSTM	0.9298	0.9129	0.9006	0.8977	0.8946
	1D-CNN	0.9359	0.9171	0.9091	0.9102	0.8917
	GRU	0.9323	0.9278	0.9054	0.9167	0.9068
Recall	LSTM	0.9170	0.9242	0.9057	0.8959	0.8926
	1D-CNN	0.9233	0.9232	0.9120	0.8919	0.9127
	DNN	0.9342	0.9295	0.9105	0.9065	0.9085
	LSTM	0.9233	0.9185	0.9032	0.8968	0.8936
F1-Score	1D-CNN	0.9295	0.9201	0.9105	0.9009	0.9021

Table 6. Deep Learning classifier Performance on SS_500 Dataset

Matrix	Model	Eyes Open	Eyes Closed	Waving left hand	Waving right hand	Mental arithmetic
Precision	DNN	0.3881	0.4072	0.3867	0.4105	0.3957
	LSTM	0.5885	0.6271	0.5330	0.5745	0.5913
	1D-CNN	0.1743	0.0	0.3684	0.3750	0.1887
	GRU	0.5598	0.4366	0.4943	0.5629	0.4882
	MS-CNN	0.3393	0.4345	0.3778	0.4742	0.5000
Recall	DNN	0.3842	0.3383	0.4028	0.3842	0.4789
	LSTM	0.7044	0.5522	0.4861	0.5320	0.6474
	1D-CNN	0.2808	0.0	0.0972	0.0148	0.6158
	GRU	0.5764	0.5821	0.3981	0.4187	0.5421
	MS-CNN	0.0936	0.7761	0.1574	0.7241	0.5211
F1-Score	DNN	0.3861	0.3696	0.3946	0.3969	0.4333
	LSTM	0.6413	0.5873	0.5085	0.5524	0.6181
	1D-CNN	0.2151	0.0	0.1538	0.0284	0.2889
	GRU	0.5680	0.4989	0.4410	0.4802	0.5137
	MS-CNN	0.1467	0.5571	0.2222	0.5731	0.5103

Table 7. Deep Learning classifier Performance on SS_250 Dataset

Matrix	Model	Eyes Open	Eyes Closed	Waving left hand	Waving right hand	Mental arithmetic
Precision	DNN	0.4956	0.5604	0.4780	0.4499	0.4163
	LSTM	0.7591	0.7634	0.6313	0.6537	0.6436
	1D-CNN	0.5763	0.5439	0.5139	0.5275	0.5734
	GRU	0.7684	0.7175	0.5990	0.6547	0.6456
	MS-CNN	0.3730	0.3658	0.4094	0.5142	0.0
Recall	DNN	0.5580	0.5062	0.4804	0.4300	0.4204
	LSTM	0.7235	0.7444	0.6127	0.6216	0.7413
	1D-CNN	0.4198	0.6154	0.4975	0.5651	0.6318
	GRU	0.6963	0.7816	0.6078	0.5823	0.7114
	MS-CNN	0.2864	0.8288	0.3431	0.5799	0.0
F1-Score	DNN	0.5250	0.5319	0.4792	0.4397	0.4183
	LSTM	0.7408	0.7538	0.6219	0.6373	0.6890
	1D-CNN	0.4857	0.5774	0.5056	0.5457	0.6012
	GRU	0.7306	0.7482	0.6034	0.6164	0.6769
	MS-CNN	0.3240	0.5076	0.3733	0.5450	0.0

Table 5는 RS Dataset에 대한 DL 분류기의 성능을 나타낸다. 대체로 DNN이 Precision, Recall, F1-Score에서 높은 성능을 보이며, RS Dataset에서는 전반적으로 Merge Dataset보다 평균 0.01 향상된 것을 보인다. 이는 비연속적인 데이터를 사용하는 분류 및 학습에서 Random Shuffling의 효과가 있을 수 있다는 것을 나타낸다.

Table 6은 SS_500 Dataset에 대한 DL 분류기의 성능을 나타낸다. Precision, Recall, F1-Score에서 LSTM이 우수한 성능을 보였다.

Table 7은 SS_250 Dataset에 대한 DL 분류기의 성능을 나타낸다. LSTM이 평가 지표에서 대부분 가장 높은 성능을 보이는 것으로 확인하였다. 또한 GRU가 LSTM보다 떨어지지만 SS_500에 비해 분류 성능이 좋아지는 것을 확인하였다. SS_500 Dataset과 SS_250 Dataset을 비교해보면 Slicing된 수가 낮아질수록 분류기의 성능이 향상되는 것을 확인할 수 있다.

RNN 계열의 분류기인 LSTM과 GRU가 연속적인 시계열 데이터를 활용하는 분류 및 학습에서 뛰어난 성능을 보여주고 있다.

Table 8. Deep Learning classifier Performance on SS_125 Dataset

Matrix	Model	Eyes Open	Eyes Closed	Waving left hand	Waving right hand	Mental arithmetic
Precision	DNN	0.5056	0.5291	0.5677	0.5066	0.5195
	LSTM	0.8003	0.8474	0.7211	0.6861	0.7296
	1D-CNN	0.7228	0.6374	0.6744	0.6270	0.5725
	GRU	0.8242	0.8168	0.7643	0.7614	0.8396
	MS-CNN	0.1936	0.0	0.0	0.0	0.0
Recall	DNN	0.6849	0.4323	0.5566	0.4885	0.4690
	LSTM	0.7972	0.7648	0.6833	0.7679	0.7606
	1D-CNN	0.5089	0.7577	0.5676	0.7090	0.6428
	GRU	0.8431	0.8682	0.7625	0.7487	0.7825
	MS-CNN	1.0	0.0	0.0	0.0	0.0
F1-Score	DNN	0.5818	0.4758	0.5621	0.4974	0.4930
	LSTM	0.7987	0.8040	0.7017	0.7247	0.7448
	1D-CNN	0.5973	0.6923	0.6164	0.6655	0.6056
	GRU	0.8335	0.8417	0.7634	0.7550	0.8101
	MS-CNN	0.3244	0.0	0.0	0.0	0.0

Table 9. Deep Learning classifier Performance on SS_25 Dataset

Matrix	Model	Eyes Open	Eyes Closed	Waving left hand	Waving right hand	Mental arithmetic
Precision	DNN	0.9343	0.9235	0.7932	0.8125	0.7684
	LSTM	0.8996	0.9429	0.8537	0.7975	0.8785
	1D-CNN	0.8932	0.8729	0.7666	0.7635	0.8483
	GRU	0.8613	0.8715	0.8332	0.7775	0.8544
	MS-CNN	0.6229	0.6933	0.5656	0.5659	0.5996
Recall	DNN	0.8808	0.8921	0.7998	0.7708	0.8729
	LSTM	0.9161	0.9038	0.7970	0.8754	0.8702
	1D-CNN	0.8576	0.8984	0.7557	0.7862	0.8429
	GRU	0.8828	0.9004	0.7572	0.8180	0.8362
	MS-CNN	0.7374	0.5281	0.3596	0.7100	0.6974
F1-Score	DNN	0.9067	0.9075	0.7965	0.7911	0.8174
	LSTM	0.9078	0.9229	0.8244	0.8346	0.8743
	1D-CNN	0.8751	0.8855	0.7611	0.7747	0.8456
	GRU	0.8719	0.8857	0.7934	0.7972	0.8452
	MS-CNN	0.6753	0.5995	0.4396	0.6298	0.6448

Table 8은 SS_125 Dataset에 대한 DL 분류기 성능을 보여준다. Precision, Recall, F1-Score에서 GRU와 LSTM 분류기가 뛰어나며, 대부분의 행동 분류에서 GRU가 우수한 성능을 보인다. SS_125 Dataset은 SS_500 및 SS_250 Dataset과 비교했을 때 Slicing 데이터가 더 적다. 이전 SS_250에서 Slicing 데이터가 감소할수록 성능이 향상되는 것을 확인했는데, SS_125 Dataset에서도 성능이 더욱 향상된 것을 확인할 수 있었다.

Table 9는 SS_25 Dataset의 DL 분류기 성능을 보여준다. LSTM과 DNN 분류기는 Precision, Recall, F1-Score에서 높은 성능을 보인다. LSTM이 대부분의 생각 및 행동 분류에서 가장 뛰어나며, SS_125 Dataset과 비교했을 때 GRU 대신 LSTM 분류기가 더 효과적인 것으로 나타났다. 시계열 데이터의 Slicing이 25로 줄어들면서 연속성이 감소하나, 여전히 RNN 계열 모델이 DNN 분류기보다 우수한 성능을 보이고 있다.

Table 10. Deep Learning classifier Performance by Dataset

Matrix	Model	Merge	RS	SS_500	SS_250	SS_125	SS_25
Precision	DNN	0.9170	0.9179	0.9179	0.4800	0.5257	0.8464
	LSTM	0.9073	0.9071	0.5829	0.6902	0.7569	0.8744
	1D-CNN	0.9108	0.9128	0.2213	0.5470	0.6468	0.8289
	GRU	-	-	0.5083	0.6771	0.8013	0.8396
	MS-CNN	-	-	0.4252	0.3325	0.0387	0.6095
Recall	DNN	0.9170	0.9178	0.3977	0.4790	0.5263	0.8433
	LSTM	0.9070	0.9071	0.5844	0.6887	0.7548	0.8725
	1D-CNN	0.9108	0.9126	0.2017	0.5459	0.6372	0.8282
	GRU	-	-	0.5035	0.6759	0.8010	0.8389
	MS-CNN	-	-	0.4545	0.4076	0.2000	0.6065
F1-Score	DNN	0.9170	0.9178	0.3964	0.4788	0.5220	0.8439
	LSTM	0.9070	0.9071	0.5815	0.6886	0.7548	0.8728
	1D-CNN	0.9107	0.9126	0.1373	0.5431	0.6354	0.8284
	GRU	-	-	0.5004	0.6751	0.8007	0.8387
	MS-CNN	-	-	0.4019	0.3500	0.0649	0.5978
Accuracy	DNN	0.9171	0.9178	0.3968	0.4790	0.5247	0.8434
	LSTM	0.9070	0.9071	0.5824	0.6884	0.7543	0.8728
	1D-CNN	0.9108	0.9126	0.1955	0.5457	0.6383	0.8285
	GRU	-	-	0.5015	0.6756	0.8015	0.8393
	MS-CNN	-	-	0.4492	0.4079	0.1936	0.6068

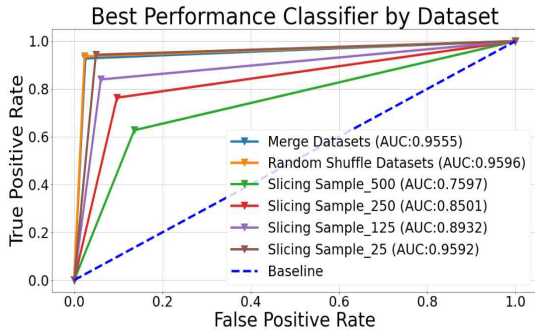


Fig. 1. Best Performance Classifier ROC Curve by Dataset

Fig. 1은 각 Dataset에 대한 생각 및 행동 분류에서 최고 성능을 보인 분류기의 ROC Curve를 보여주며, 이는 Random Shuffle Dataset이 분류 하는데 있어 가장 적합한 Dataset임을 보여준다. Table 10은 Accuracy Matrix와 함께 성능을 나타낸다. DNN은 연속성을 고려하지 않은 경우에 강한 성능을 보이고, LSTM은 연속성이 있는 데이터에서 우수한 성능을 나타냈다. 이로써 연속성 여부에 따라 DNN과 RNN 계열의 분류기가 각각 우세함을 확인할 수 있었다.

IV. 결론

본 연구는 BCI(뇌-컴퓨터 인터페이스) 연구에서 데이터의 구성에 따른 다양한 모델의 성능을 비교하기 위한 실험을 수행하였다. 특히, 시계열 데이터의 연속성에 주목하여, 5가지 분류로 수집된 EEG 신호를 연속성에 따라 6개의 그룹으로 나누어 실험을 진행하였다. 실험에서는 DNN, LSTM, 1D-CNN과 같은 분류 모델을 사용하였으며, 데이터의 연속성을 고려한 슬라이싱 서플 데이터에서는 GRU 및 Multi-Scale CNN 모델도 추가로 실험하였다. 실험 결과를 통해, 데이터의 연속성을 고려하지 않은 경우 DNN 분류기가 강력한 성능을 보였으며, 특히 Merge Dataset과 RS Dataset에서 우수한 결과를 나타냈다. 또한, RNN 계열의 분류기는 연속성이 있는 분류(SS_Dataset)에서 일반적으로 높은 성능을 보였다. 특히, 연속성이 있는 SS_500, SS_250, SS_25 Dataset에서 LSTM 모델이 뛰어난 성능을 보였다. 그러나, SS_125 Dataset에서는 GRU 분류기가 미세한 차이로 최고의 성능을 보였다. 이러한 결과를 종합적으로 살펴보면, EEG 신호를 기반으로 하는 BCI 연구에서는 데이터의 연속성이 적을수록 딥러닝 분류기의 성능이

향상되는 경향이 있음을 확인할 수 있다. 특히, RNN 계열 중 LSTM이 데이터의 연속성을 고려할 때 높은 성능을 보였으며, 특정 범위로 연속된 데이터에서는 간소화된 계산방식을 가진 GRU가 강점이 될 수 있다고 보았다. 데이터의 연속성을 고려하지 않는 경우에는 DNN이 가장 높은 성능을 보인다. 또한, 무작위 서플링을 수행한 데이터가 전체적으로 가장 높은 성능을 나타내는 결과를 얻었는데, 이는 연속성을 고려하지 않는 데이터가 분류에 더 효과적인 것을 시사한다. 이에 따라, EEG 신호를 활용한 BCI 연구에서는 데이터의 연속성을 고려하기보다는 서플링을 통한 데이터 특징 다변화가 분류 성능 향상에 기여할 수 있다는 결론을 도출하였다. 이 연구 결과는 BCI 연구 분야에서 데이터 구성에 따른 연구 방향에 긍정적인 기여가 있을 것으로 생각한다. 실시간성이 보장이 되는 BCI 연구에서 행동을 정확하게 유지하는데 연속된 행동에서의 데이터 연속성이 분류에 있어서 어떠한 영향이 가는지에 대한 도움이 될 것으로 예상된다. 추후 연구에서는 파라미터에 의한 발생될 수 있는 차이를 다양한 방면에서의 반복적인 실험으로 오차범위를 줄여 나가는 연구를 진행할 것이며, 실시간성을 보장하는 뇌파 분류 연구를 진행할 예정이다.

ACKNOWLEDGMENTS

이 논문은 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 ‘바이오융복합기술 전문인력 양성사업’의 지원을 받아 수행된 연구임(No. P0017805).

REFERENCES

- [1] C. G. Langton, “Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, Held September, 1987, in Los Alamos, New Mexico,” in *Artificial Life*, xth ed., Addison-Wesley Pub. Co., Advanced Book Program, 1989.
- [2] G. Macaudo, “Artifacts removal and application of cross-frequency coupling to EEG signals of epileptic patients,” Webthesis, 2020.
- [3] W.-S. Choi and H.-G. Yeom, “Studies to overcome brain-computer interface challenges,” *MDPI*, Vol.12(5), pp.2598, 2022.

- [4] M. Lo Giudice et al., "1D convolutional neural network approach to classify voluntary eye blinks in EEG signals for BCI applications," in Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), 2020, doi:10.1109/ijcnn48605.2020.9207195.
- [5] S. Cheng, J. Wang, L. Zhang, and Q. Wei, "Motion imagery-BCI based on EEG and eye movement data fusion," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 28, no. 12, pp. 2783-2793, 2020, doi:10.1109/tnsre.2020.3048422.
- [6] H. S. Lee et al., "Individual finger movement decoding using a novel ultra-high-density electroencephalography-based brain-computer interface system," Front. Neurosci, vol.16, 2022.
- [7] J. Shin, J. Kwon, and C.-H. Im, "A ternary hybrid EEG-NIRS brain-computer interface for the classification of brain activation patterns during mental arithmetic, motor imagery, and Idle State," Front. Neuroinformatics, vol.12, 2018, doi:10.3389/fninf.2018.00005.
- [8] R. Chai et al., "Selecting optimal EEG channels for mental tasks classification: An approach using ICA," in Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), 2016, doi:10.1109/cec.2016.7743942.
- [9] R. Darmakusuma et al., "Analysis of arm movement prediction by using the Electroencephalography signal," Makara Journal of Technology, vol. 20, no. 1, pp.38, 2016, doi:10.7454/mst.v20i1.3054.
- [10] C. Huang, Y. Xiao, and G. Xu, "Predicting human intention-behavior through EEG signal analysis using multi-scale CNN," IEEE/ACM Trans. Comput. Biol. Bioinformatics, vol.18, no.5, pp.1722-1729, 2021, doi:10.1109/tcbb.2020.3039834.
- [11] J.-H. Cho, J.-R. Jeong, D.-J. Kim, and S.-W. Lee, "A novel approach to classify natural grasp actions by estimating muscle activity patterns from EEG signals," in Proceedings of the 2020 8th International Winter Conference on Brain-Computer Interface (BCI), 2020, doi:10.1109/bci48061.2020.9061627.

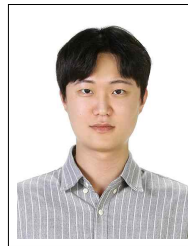
저자소개

김 윤 상 (Youn-Sang Kim)



2021년 2월 : 건양대학교
의료IT공학과(공학사)
2022년 9월~현재 : 건양대학교
의료공학과(공학석사)
관심분야 : 인공지능, 의료공학,
뇌과

한 주 혁 (Ju-Hyuck Han)



2018년 8월 : 건양대학교
의료IT공학과(공학사)
2020년 8월 : 건양대학교
의료공학과(공학석사)
2021년 3월~현재 : 건양대학교
의료공학과(공학박사)
관심분야 : 인공지능, 의료공학,
전자공학

김 웅 식 (Sik-Woong Kim)



1989년 2월 : 인하대학교
정보공학과(공학석사)
2007년 2월 : 인하대학교
컴퓨터공학과(공학박사)
2006년 3월~현재 : 건양대학교
인공지능학과 교수
관심분야 : 인공지능, 의료공학,
임베디드, 뇌과