Comparison experiment of artificial intelligence models based on EEG wave continuity

Kim Youn-Sang
Department of Medical Engineering
Konyang University
Daejeon, South Korea
yskim2570@gmail.com

Han Ju-Hyuck
Department of Medical Engineering
Konyang University
Daejeon, South Korea
dnfwlq203@gmail.com

Lee Han-Jo
Representative Director
Safecoms Co.
Seoul, South Korea
leehjgo@naver.com

Kim Woong-Sik*
Department of Artificial Intelligence
Konyang University
Daejeon, South Korea
wskim@konyang.ac.kr

***Abstract*— In this study, we conducted comparative experiments on continuous data for BCI research using various comparison models depending on the data composition. The goal was to compare the classification performance of artificial intelligence models on data that considered continuity, a characteristic of time series data, and data that did not. To summarize the research results in BCI research using EEG signals, action and thought classification achieved a high-performance score of 0.8728 in LSTM as data continuity decreased. Additionally, DNN showed maximum performance with a score of 0.9178 when continuity was not considered. Moreover, it was observed that the data without considering continuity performed well in task classification. Therefore, BCI research focusing on behavior and accident classification based on EEG signals is expected to achieve excellent performance by showing various data characteristics through shuffling rather than considering data continuity.**

***Keywords*— EEG signals, BCI research, Time series data, Classification Performance, DNN, Data continuity**

I. INTRODUCTION

EEG waves are measurable electrical signals generated by neurons as weak electrical signals propagate through tens of thousands of neurons [1]. Such EEG patterns can be categorized largely into five types based on their frequency. Delta (δ) waves have a frequency of 0.2 to 4 Hz and occur during very deep sleep, meditation, or unconsciousness. Theta (θ) waves have a frequency of 4 to 8 Hz and are associated with creative thinking, emotional stability, and falling asleep. Alpha (α) waves have a frequency of 8 to 13 Hz and occur when the mind is relaxed and calm. Stable alpha waves are produced when the eyes are closed and relaxed, and these are affected by the opening and closing of the eyes. Because of this, it is speculated that alpha waves are associated with the visual areas of the brain. Mu (μ) waves, a subset of alpha waves, are observed at frequencies of 7.5 to 12.5 Hz (mostly 9 to 11) and are best observed when the body is in physical motion. Beta (β) waves with a frequency of 13 to 30 Hz are generated during everyday cognitive and thinking activities. It is often generated when doing complex calculations or thinking. Gamma (γ) waves with a frequency of 30 Hz or higher are produced when a person is in a state of high concentration, such as extreme tension or excitement [2]. A brain-computer interface (BCI) is a technology that measures the EEG signals from the brain to control a computer or machine using only thoughts [3]. Research on utilizing these BCIs is ongoing in a variety of areas. EEG is a time series of signal data with continuously changing potentials. According to the signal characteristics of EEG, it is classified into time series, oscillation, topographic information, and net structure analysis, among which time series analysis shows the continuous change of potential over time [4]. Time series data is a sequence of data points placed at regular time intervals. Time series data has a characteristic that it is continuous over time. This study focused on the temporal continuity of these EEG data.

TABLE I. COMPARISON OF EXISTING STUDIES

Author	Dataset	Preprocessing	Using Model
Lo Giudice, et al.[5]	Manual	Butterworth filter	1D-CNN
Cheng, S., et al.[6]	Manual	0.5Hz Cut-off Hz. Wavelet Transform	SVM
Khan, S.U., et al.[7]	Public	8th Chebyshev, 50Hz NF, STFT	VGG-11, 13, 16, 19, FBCSP+CNN
Lee, H.S., et al.[8]	Manual	NF, 4th Butter worth filter.	Linear SVM
Ieracitano, C., et al.[9]	Public	CWT.	Hybrid-domain DL approach
Planelles, D., et al.[10]	Manual	50Hz NF, 8 Butterworth filter, Spatial filter, Laplacian algorithm.	SVM
Kline, A., et al.[11]	Manual	2nd 0 BPF, DFT	Multi-Perceptron
Bose, R., et al.[12]	Manual	PSD,Statistical parameters, Hjorth parameter	kNN, SVM
Kwon, M., et al.[13]	Manual	FFT, BPF	MI-BCI classification
Shin, J., et al.[14]	Manual	0.1~50Hz BPF	sLDA
Yavuz, E., et al.[15]	Public	CSP	kNN, LDA, NB, DT, SVM, RF
Rahman, M.M., et al.[16]	Manual	0.5Hz HPF, 50Hz NF, 30Hz LPF	RTA-NET
Chai, R., et al.[17]	Manual	PSD, AR Modeling	MLP
Darmakusuma, R., et al.[18]	Manual	Butterworth BPF	SVM
Huang, C., et al.[19]	Public	0.5 ~ 100Hz BPF, STFT	ANN,SVM,SAE, MS-CNN
Cho, J. H., et al.[20]	Manual	60Hz NF	CSP, LDA Model
Venkatachalam, K., et al.[21]	Public	4 ~ 40Hz BPF	ELM, Kernel ELM, HELM, KHELM, Hybrid-KELM

Note.—Manual = Manual Annotation, Public = Public Dataset, NF = Notch Filter, STFT = Short Time Fourier Transform, CWT = Continuous Wavelet Transform, BPF = Band-Pass Filter, DFT = Discrete Fourier Transform, PSD = Power Spectral Density, FFT = Fast Fourier Transform, CSP : Common Spatial Pattern, AR = Autoregressive, 1D-CNN = 1-dimensional Convolutional Neural Network, SVM = Support Vector Machine, VGG = Visual Geometry Group, FBCSP = Filter Bank Common Spatial Pattern, CNN = Convolutional Neural Network, DL = Deep Learning, kNN = k-Nearest Neighbor, MI-BCI = Motor Imagery-Based Barin-Computer Interface, sLDA = supervised Linear Discrimination Analysis, LDA = Linear Discrimination Analysis, NB = Naïve Bayes, DT = Decision Tree, RF= Random Forest, RTA-NET = Residual-based Temporal Attention Network, MLP = Multi-Layer Perceptron, ANN = Artificial Neural Network, SAE = Stacked Auto-Encoder, MS-CNN = Multi-Scale Convolutional Neural Network, ELM = Extreme Learning Machines, HELM = Hybrid Extreme Learning Machines, KHELM = Kernel Hierarchical Extreme Learning Machine, KELM = Kernel Extreme Learning Machine

The behaviors in this study were selected through a comparative analysis of what brainwaves were used in previous studies. After selecting behaviors, an AI model was selected for this study. This study selected the 1D-CNN used by Lo Giudice et al.; DT by Yavuz, E., et al.; MS-CNN and ANN (DNN) by Huang, C., et al.; and LSTM and GRU models that can utilize continuity, which is a characteristic of time series. As many existing EEG studies had difficulty obtaining and using public data for their research, they acquired data through manual annotation. It is difficult to obtain public data because there are issues about the anonymization and de-identification of EEG data as human-derived signals, as well as inconveniences in protecting subjects and disclosing data. When conducting new research, there is no data that meets the research purpose. Therefore, new data is acquired directly to conduct research. However, there are difficulties in disclosing the data since it is also human-derived. Many of these EEG signal studies use different datasets and utilize different models due to the difficulty of obtaining public data, as mentioned above. Other studies have considered continuity when utilizing EEG data but have not been quantitatively consistent in their approach. Due to data acquisition constraints, some studies have acquired their own EEGs, while others have used public data. However, the data descriptions of these studies are not uniform in their consideration of data preprocessing or continuity, making it difficult to study under the same conditions as previous studies. For these reasons, this study is organized by preprocessing the data and organizing the data in a way that unifies continuity. Previous studies had difficulties comparing the performance of each model. To address these difficulties, a study was conducted to organize a comparison experiment of BCI research based on EEG signals. Some of the behaviors and models used in other existing studies were selected for this study. The study was conducted by dividing the data structure into three methods according to the continuity of the time series of EEG data. Moreover, comparative analysis was performed on six AI models. This study aims to determine the impact of continuity in time series data on EEG classification performance and identify optimal classification models for different datasets. The results of the study are expected to provide useful information for other BCI research.

II. METHODS

A. Datasets

In this study, an experiment was conducted for the comparative analysis of BCI research based on EEG signals. EEG data was collected from real human participants using a non-invasive EEG device (Quick-20r from CGX A cognitionics company). The purpose and process of the experiment were explained to all participants, and they signed an informed consent form. The study was approved by the Institutional Review Board (IRB) of Konyang University (KYU 2022-12-004-001). The electrodes used to collect the EEG data were based on the international 10-20 system established by the International Federation of Societies for Electroencephalography and Clinical Neurophysiology [22]. The data collection process consists of three steps as follows: ① Place the CGX Quick-20r on the participant's head, and adjust the electrodes for each electrode position. ② Check whether EEG is read normally through Bluetooth communication between Quick-20r and the laptop. ③ Participants perform each behavior for 30 seconds to collect EEG signals. For mechanical classification (annotation) of behaviors, this dataset consists of a similar amount of data for each of the five behaviors according to the methods used in existing studies: eyes open and eyes closed used by Lo Giudice et al.; waving left hand and waving right hand based on hand movements used by Khan et al. and Planelles et al.; and a mental arithmetic condition used by Shin et al. These are classifications of behaviors and thoughts, and EEG measurements are taken for 30 seconds per classification, resulting in two and a half minutes of signals from one participant. The configuration for data classification is as follows:

- Eyes Open
- Eyes Closed
- Waving Left Hand
- Waving Right Hand
- Mental Arithmetic

B. Preprocessing

This study preprocesses the dataset obtained through data collection. Sampling rate is 500 Hz, and the EEG channel is 21Ch. The acquired EEG data was measured at 0.002 second intervals for 30 seconds per behavior. The data was filtered using a band-pass filter.

C. Experiment

In this study, the experimental dataset is divided into three groups to evaluate the data, considering the continuity of time series data according to the data configuration. Table 2 shows the data constructed for the experiments in this study, and the data is categorized into six classifications based on continuous range and random shuffle.

TABLE II. CONFIGURATION OF DATASETS

Factor \ Dataset	Merge	RS†1	SS†2_500	SS_250	SS_125	SS_25
*CR (sec)	150	0	1	0.5	0.25	0.05
**RS (True/False)	F	T	T	T	T	T

Note.— *CR : Continuous Range, **RS : Random Shuffling, RS†1 : Random Sample, SS†2 : Slicing Sample.

The first is the merge data, where all the consecutive data are consistently consolidated. This is where the continuity of the data after preprocessing is used as a feature in the classification model, and the data is a time series at the time of measurement. The second is the random sample (RS) dataset, which is typically used to avoid overfitting and bias, which are problems with statistical classification methods when training a classification model. The RS data is a big limitation to data continuity, but it is used in many AI studies to solve problems with statistical classification methods. Where, all data are shuffled randomly to avoid considering continuity. This experimental data shows how continuous data like EEG affects classifiers based on its nature. The third is the slicing sample (SS) data, which has been truncated to account for continuity and partially randomly shuffled to account for continuity. The SS data is categorized into four types (1s, 0.5s, 0.25s, and 0.05s) based on the number of seconds over which continuity is guaranteed. In this study, experiments with AI classification models are conducted on six classified datasets. The AI models used in the classification experiments have been selected from EEG-related studies. The models are DT, DNN, LSTM, 1D-CNN, GRU, and MS-CNN, and the classification experiments are conducted using DT, DNN, 1D-CNN, GRU, and MS-CNN on the Merge dataset and RS dataset. For experiments with the SS dataset, which considers continuity, GRU and multi-scale CNN models have been added. GRU and MS-CNN were used for the SS dataset to check the strength of existing LSTM, 1D-CNN, and improved AI models in classifying data considering continuity. To keep the comparative classification experiments in the same research environment as much as possible, the complexity of the classification models, the hierarchy, and the optional variables were all configured similarly. This enables a comparison of how the structure of each classification model might perform, depending on the data.

TABLE III. TEST MODEL CONFIGURATION BY DATASETS

Dataset / Model	Merge	RS	SS_500	SS_250	SS_125	SS_25
DT	Depth 25	Depth 25	Depth 25	Depth 25	Depth 25	Depth 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	291,549	291,549	293,817	292,407	291,957	291,957
GRU†3	-	-	293,561	292,901	292,521	292,245
MS-CNN	-	-	299,175	299,385	290,845	294,205

Note.— DNN†1 : Deep Neural Network, LSTM†2 : Long Short Term Memory, GRU†3 : Gated Recurrent Unit

Table 3 shows the complexity of the classification models in the comparison experiments on the dataset. The hyper-parameters of all experiments in this study are set to the same conditions. The set conditions were that the optimizer used Adam Optimizer and the batch size was set to 32 and epochs 50. The deep learning (DL) classifiers have a similar complexity of approximately 300,000 except for DT, which is a machine learning (ML) classifier. In the DT classifier, the depth of the tree is the main parameter that determines the performance. In this study, it was increased from 1 to 24 to check the performance. For DL classifiers, the hierarchy varies depending on the division of the dataset being fed into the model. The different layers cause the slight differences in complexity between the models, as shown in Table 2, and the goal is to feed the classification model without compromising the continuity of the dataset. In addition, all classification models were set up based on a four-layer configuration. However, for models that require a large number of layers to represent the performance of the classification model, such as CNN and LSTM, only a minimal number of layers were added. Along with the CNN family, an improved classifier from the recurrent neural network (RNN) family, which is utilized in many studies in DL classifiers, was added. RNN classification models, such as LSTM and GRU, are known for their strengths in sequential data by storing state vectors of features to influence the next vector. GRU and MS- CNN, which were newly added to the comparison experiments, were not tested on the merge and RS datasets. To examine and analyze the strength of GRU and MS- CNN as improved forms of AI models for classifying data considering continuity in detail, they were tested with the SS dataset and not with the merge and RS datasets.

III. RESULT

This study conducted experiments on how continuous features perform across classification models, depending on the data configuration. The experimental data was divided into three groups, and representative ML and DL classifiers were selected and compared for the BCI classification problem.

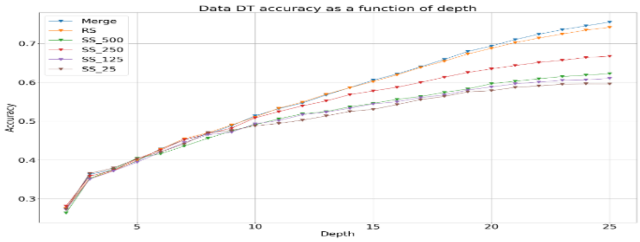

Fig. 1. Decision Tree Classifier Performance Graph

TABLE IV. DEEP LEARNING CLASSIFIER PERFORMANCE ON MERGE DATASET

Matrix	Models	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 4 shows the performance results of the DL classifiers with the merge dataset. In terms of the precision of classification, the DNN classifier showed good performance in general. Only for eyes open, the LSTM classifier showed the best classification performance. In terms of recall, the DNN classifier performed best for eyes open and closed as well as waving right hand. The 1D-CNN classifier performed best for recall of the waving left hand, while the LSTM classifier performed best for mental arithmetic. Finally, the DNN classifier performed the best for all classifications in terms of F1-Score.

TABLE V. DEEP LEARNING CLASSIFIER PERFORMANCE ON RS DATASET

Matrix	Models	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
Recall	DNN	0.9323	0.9278	0.9054	0.9167	0.9068
	LSTM	0.9170	0.9242	0.9057	0.8959	0.8926
	1D-CNN	0.9233	0.9232	0.9120	0.8919	0.9127
F1-Score	DNN	0.9342	0.9295	0.9105	0.9065	0.9085
	LSTM	0.9233	0.9185	0.9032	0.8968	0.8936
	1D-CNN	0.9295	0.9201	0.9105	0.9009	0.9021

Table 5 shows the performance results of the DL classifiers with the RS dataset. As shown in the table, DNNs showed outstanding performance for most classifications in terms of precision, recall, and F1-Score. With the merge data, the highest classification performance was seen in precision for the classification of eyes open and recall for the classification of mental arithmetic. The classification performance for eyes open was increased with the RS dataset, but it did not achieve good results as the classification performance of DNN was also increased. Moreover, the overall evaluation matrix was 0.01 higher than with the merge dataset on average with the RS dataset. This suggests the effectiveness of random shuffling in classification and learning that utilizes discrete data.

TABLE VI. DEEP LEARNING CLASSIFIER PERFORMANCE ON SS_500 DATASET

Matrix	Models	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 VII. DEEP LEARNING CLASSIFIER PERFORMANCE ON SS_250 DATASET

Matrix	Models	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

Tables 6 and 7 show the performance of the DL classifiers with the SS_500 and SS_250 datasets, respectively. LSTM and GRU outperformed in terms of precision, recall, and F1-score. Comparing the SS_500 and 250 datasets, the performance of the classifiers improved as the number of slices decreased. LSTM and GRU, which were classifiers of the RNN family, outperformed when it came to classification and learning effectiveness using continuous time series data.

TABLE VIII. DEEP LEARNING CLASSIFIER PERFORMANCE ON SS_125 DATASET

Matrix	Models	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 8 shows the performance results of the DL classifiers with the SS_125 dataset. The GRU and LSTM classifiers performed well in terms of precision, recall, and F1-score. For the SS_125 dataset, the GRU classifier performed best for most thought and behavior classifications. The SS_125 dataset had less sliced data compared to the SS_500 dataset and SS_250 dataset above. Previously, the performance of the performance evaluation matrix increased with SS_250 as the slicing data decreased, and the performance of the SS_125 dataset was also better.

TABLE IX. DEEP LEARNING CLASSIFIER PERFORMANCE ON SS_25 DATASET

Matrix	Models	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 9 shows the performance results of the DL classifiers with the SS_25 dataset. LSTM and DNN outperformed in terms of precision, recall, and F1-score. The LSTM classifier performed best for most thought and behavior classifications. Previously, the GRU classifier performed well in the SS_125 dataset; however, the LSTM classifier performed better in the SS_25 dataset,. As the slicing of the time series data was reduced to about 25, the continuity of the time series data was greatly reduced. Moreover, the DNN model, which previously showed excellent performance, may have shown good performance in classifying time series data. However, the RNN family models seemed to outperform the DNN classifier in classification performance with time series data.

TABLE X. DEEP LEARNING CLASSIFIER PERFORMANCE BY DATASET

Matrix	Models	Merge	RS	SS_500	SS_250	SS_125	SS_25
Precision	DNN	0.9170	0.9179	0.3976	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	0.3961	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

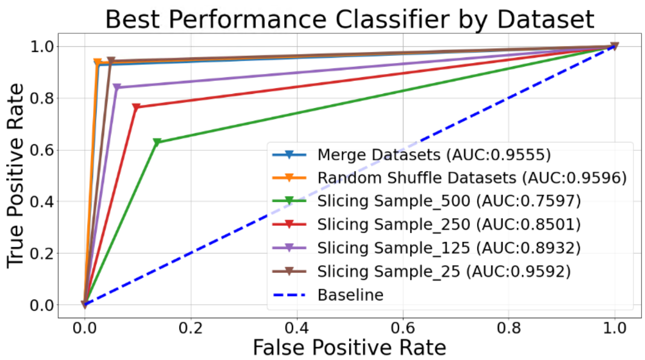

Fig. 2. Best Performance Classifier ROC Curve by Dataset

Figure 2 shows the receiver operating characteristic (ROC) curve of the best-performing classifier for thought and behavior classification based on the dataset. The best model for each dataset from the classification performance metrics was selected in Tables 4 to 9 to provide a per-classification area under the curve (AUC). This included the average AUC of the five classification models. Table 10 is different from the other tables of performance results because of the addition of the accuracy matrix. As for the accuracy according to the data, DNN showed the best classification performance with the merge and RS datasets, and LSTM showed the best classification performance with all SS_datasets except 125. In the SS_125 dataset, the GRU classifier achieved the best performance. Experiments on classifying thoughts and behaviors based on data continuity showed that DNN classifiers performed strongly when data continuity was not considered. It performed best with the merge and RS datasets. The 1D-CNN and MS-CNN classifiers of the CNN family performed worse than the RNN family on continuous classifications like the SS datasets. On the other hand, the RNN family showed strong classification performance when there is continuity in the data, and the LSTM model had the best classification performance with the SS_500, SS_250, and SS_25 datasets. However, the GRU classifier outperformed the LSTM in the SS_125 dataset. In summary, the results of this study suggest that DNN classifiers perform well when data continuity is not considered, and RNN-based classifiers perform well when continuity is considered.

IV. CONCLUSION

In this study, a comparative experiment was conducted on data continuity for BCI research with different comparison models based on data organization. This aimed to compare the classification performance of each model under the same conditions of data composition and continuity consideration, which were characteristics of time series data. This study focused on the continuity of the time series data and used six EEG signals collected in five different classifications based on their continuity. The experiments evaluated classification performance by training AI models of similar complexity with classified data. The classification models for comparison experiments were DT, DNN, LSTM, 1D-CNN, GRU, and MS- CNN models were further experimented with the SS data considering data continuity. Experimental results showed that DNN classifiers performed strongly when data continuity was not considered. The DNN classifier performed best with the merge and RS datasets. RNN classifiers also performed well on average for classifications with continuity, such as the SS_datasets. We paid attention to this. Across the SS_500, SS_250, and SS_25 datasets, with continuity in the data, the LSTM model performed the best across all evaluation matrices. The GRU classifier showed the best performance in the SS_125 dataset, with a slight performance difference compared to LSTM. Upon reviewing the experimental results of this study, it was found that lesser data continuity has better performance of deep learning classifiers in BCI research based on EEG signals. The LSTM was expected to achieve the best performance across the SS_datasets, since the performance of the LSTM was consistently stronger with the SS_125 dataset. However, the GRU classifier performed better in the SS_125 dataset, suggesting that RNN classifiers performed differently depending on the number of continuities in the data. However, the LSTM classifier, not the GRU classifier, achieved the best performance in the SS_25 dataset. Regarding this phenomenon, this study determined that the simplified computational approach of the GRU model could be a strength for data with continuity over a certain range. In summary, the behavior and thought classification were best performed by the RNN family, especially LSTM when the continuity of the data was considered and DNN when the continuity of the data was not considered in BCI research based on EEG signals. In addition, the performance of the classifier according to the dataset showed that the data with random shuffling was the best-performing dataset overall. This suggests that the dataset that does not take continuity into account is a better dataset for classification than one that does. Therefore, BCI research on behavior and thought classification based on EEG signals seems to benefit from data feature diversification through shuffling rather than ensuring data continuity. Accordingly, this study is expected to make a positive contribution to the direction of BCI research based on data continuity. In the future work will diversify feature vector with the transformer models.

ACKNOWLEDGMENT

Following are results of a study on the "Leaders in INdustry-university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea (NTIS NO. 1345370645).

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. Addison-Wesley Pub. Co., Advanced Book Program, 1989.

[2] G. Macauda, "Artifacts removal and application of cross-frequency coupling to EEG signals of epileptic patients," Webthesis, https://webthesis.biblio.polito.it/16971/

[3] W.-S. Choi and H.-G. Yeom, "Studies to overcome brain–computer interface challenges," MDPI, https://www.mdpi.com/2076-3417/12/5/2598. 2022.

[4] H. J. Eun, "Basics of electroencephalography for Neuropsychiatrist," Journal of Korean Neuropsychiatric Association, vol. 58, no. 2, p. 76, 2019. doi:10.4306/jknpa.2019.58.2.76

[5] M. Lo Giudice et al., "1D convolutional neural network approach to classify voluntary eye blinks in EEG signals for BCI applications," 2020 International Joint Conference on Neural Networks (IJCNN), 2020. doi:10.1109/ijcnn48605.2020.9207195

[6] S. Cheng, J. Wang, L. Zhang, and Q. Wei, "Motion imagery-BCI based on EEG and eye movement data fusion," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 12, pp. 2783–2793, 2020. doi:10.1109/tnsre.2020.3048422

[7] S. U. Khan, M. Majid, and S. M. Anwar, "Upper limb movement execution classification using electroencephalography for Brain Computer Interface," arXiv.org, https://arxiv.org/abs/2304.06036.

[8] H. S. Lee et al., "Individual finger movement decoding using a novel ultra-high-density electroencephalography-based brain-computer interface system," Frontiers in neuroscience, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9627315/.

[9] C. Ieracitano, F. C. Morabito, A. Hussain, and N. Mammone, "A hybrid-domain deep learning-based BCI for discriminating hand motion planning from EEG Sources," International Journal of Neural Systems, vol. 31, no. 09, p. 2150038, 2021. doi:10.1142/s0129065721500386

[10] D. Planelles et al., "Evaluating classifiers to detect arm movement intention from EEG Signals," Sensors, vol. 14, no. 10, pp. 18172–18186, 2014. doi:10.3390/s141018172

[11] A. Kline, C. Gaina Ghiroaga, D. Pittman, B. Goodyear, and J. Ronsky, "EEG differentiates left and right imagined lower limb movement," Gait & Posture, vol. 84, pp. 148–154, 2021. doi:10.1016/j.gaitpost.2020.11.014

[12] R. Bose, A. Khasnobish, S. Bhaduri, and D. N. Tibarewala, "Performance analysis of left and right lower limb movement classification from EEG," 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), 2016. doi:10.1109/spin.2016.7566683

[13] M. Kwon, H. Cho, K. Won, M. Ahn, and S. C. Jun, "Use of both eyes-open and eyes-closed resting states may yield a more robust predictor of motor imagery BCI performance," Electronics, vol. 9, no. 4, p. 690, 2020. doi:10.3390/electronics9040690

[14] 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," Frontiers in Neuroinformatics, vol. 12, 2018. doi:10.3389/fninf.2018.00005

[15] E. Yavuz and O. Aydemir, "Classification of mental arithmetic based hybrid EEG+NIRS signals," 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020. doi:10.1109/siu49456.2020.9302051

[16] M. M. Rahman and M. Aktaruzzaman, "An end-to-end deep learning model for mental arithmetic task classification from multi-channel EEG," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), 2021. doi:10.1109/becithcon54710.2021.9893638

[17] R. Chai et al., "Selecting optimal EEG channels for mental tasks classification: An approach using ICA," 2016 IEEE Congress on Evolutionary Computation (CEC), 2016. doi:10.1109/cec.2016.7743942

[18] R. Darmakusuma et al., "Analysis of arm movement prediction by using the Electroencephalography signal," Makara Journal of Technology, vol. 20, no. 1, p. 38, 2016. doi:10.7454/mst.v20i1.3054

[19] C. Huang, Y. Xiao, and G. Xu, "Predicting human intention-behavior through EEG signal analysis using multi-scale CNN," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 5, pp. 1722–1729, 2021. doi:10.1109/tcbb.2020.3039834

[20] 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," 2020 8th International Winter Conference on Brain-Computer Interface (BCI), 2020. doi:10.1109/bci48061.2020.9061627

[21] V. K. et al., "A novel method of motor imagery classification using EEG signal," Artificial Intelligence in Medicine, vol. 103, p. 101787, 2020. doi:10.1016/j.artmed.2019.101787