Avicenna Journal of Neuro Psycho Physiology
doi: 10.32592/ajnpp.2023.10.1.102 2023 February;10(1): 15-21 https://ajnpp.umsha.ac.ir

Original Article

# Deep Learning-Based Approach for Classification of Mental Tasks from Electroencephalogram Signals

**Evin Şahin Sadık[1]*, Hamdi Melih Saraoglu[1], Sibel Canbaz Kabay[2], Cahit Keskinkılıç[3]**

[1] PhD Kütahya Dumlupinar University, Faculty of Engineering Dept. of Electrical Electronics Eng., Kütahya, Turkey. Kütahya Dumlupınar University
[2] PhD Kütahya Health Sciences University, Faculty of Medicine, Dept. of Neurology, Kütahya, Turkey - Kütahya Sağlık Bilimleri Üniversitesi
[3] MsC Department of Psychology, İstanbul Gedik University, İstanbul, Turkey. Istanbul

***Corresponding author:**
Evin Şahin Sadık. Kütahya Dumlupınar University Electrical-Electronics Engineering, Turkey.
Tel:334244327490
Email: evin.sahin@dpu.edu.tr

Received: 18 Jan 2023
Accepted: 14 May 2023
ePublished: 20 Jun 2023

**Abstract**

**Background and Objective:** Electroencephalography (EEG) analysis is an important tool for neuroscience, brain-computer interface studies, and biomedical studies. The primary purpose of Brain-Computer Interface (BCI) studies is to establish communication between disabled individuals and other individuals and machines with brain signals. Interpreting and classifying the brain's response during different cognitive tasks will contribute to brain-computer interface studies. Therefore, in this study, 5 cognitive tasks were classified from EEG signals.

**Materials and Methods:** In this study, five neuropsychological tests (Öktem Verbal Memory Processes Test, WMS-R Visual Memory Subtest, Digit Span Test, Corsi Block Test and Stroop Test) were administered to 30 healthy individuals. The tests assess the volunteers' abilities in verbal memory, visual memory, attention, concentration, working memory and reaction time. EEG signals were recorded while the tests were administered to the volunteers. Tests were classified using two different deep learning algorithms, 1D Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), from the recorded EEG signals.

**Results:** When the test successes were evaluated, classification success was achieved with an accuracy of 88.53% in the CNN deep learning algorithm and 89.80% in the LSTM deep algorithm. Precision, recall and F1-score values for CNN are 0.88, 0.87 and 0.87, respectively, while precision, recall and f1-score values for the LSTM network are 0.90, 0.89 and 0.89.

**Conclusions:** Following the findings of the present study, five different cognitive tasks were able to be classified with high accuracy from EEG signals using deep learning algorithms.

**Keywords:** EEG, Cognitive task, Deep learning, Classification, Working memory

## Background

In recent years, brain-computer interface (BCI) studies that help people with physical disabilities communicate with brain Electroencephalogram (EEG) signals have been popular topics of study. BCI research aims to establish a communication system by translating human intentions, reflected by appropriate brain signals, into control signals in output devices such as a computer application or neuroprosthesis [1]. The success of BCI applications depends on the detection and classification accuracy of mental tasks [2]. Through BCI, it is possible to read brain neural activities accurately and communicate or interact with the environment without requiring muscle use or physical action. BCI studies emerge with the collaboration of experts in the fields of neurology, psychology and computer science [3]. In a successful BCI application, it is necessary to apply the science of psychology to interpret mental tasks, the science of neurology to examine the changes that occur in the brain during mental tasks, and computer science for the transfer and processing of EEG signals to the computer. However, manual examination of EEG signals by an expert, extracting features from EEG signals, and identifying features related to cognitive activities is a time and labor-consuming process. Leaving these processes in the computer-based machine learning mechanism can reduce the probability of error and ensure the effective use of spent resources [4].

When the literature is examined, it is seen that the classification of cognitive activities carried out so far is generally classified as having a task, no task, or classifying two tasks. [5–9]. In multitasking classification studies, it is seen that classification is made using mostly machine learning methods, feature extraction and feature selection methods. In a study in which visual attention/inattention states were classified from EEG signals, an accuracy of 85.24% was obtained with the support vector machines (SVM) algorithm. [5]. In a study in which the effects of the cognitive task on visual pattern recognition on the brain and the brain functions at rest were classified, SVM, K nearest neighbor (KNN) and Naive Bayesian (NB) algorithms were used and subband features obtained using the wavelet decomposition method were extracted, and 99.11 % accuracy achieved [6]. With the features obtained by the Fourier decomposition method, 98.6% success was achieved with the SVM algorithm in the classification of mental arithmetic tasks [7]. In a recent study, deep learning methods were used to classify four different mental tasks, but 88.33% classification accuracy was obtained by extracting features from frequency bands with the Cross Frequency Coupling method [4].

In this study, EEG signals were recorded while 30 volunteers were performing 5 different cognitive tasks. EEG signals recorded from 16 channels, after preprocessing, were used to classify cognitive tasks with deep learning methods. With five different cognitive task LSTM algorithms, 89.80% classified with high accuracy. Our method has a high success in multiclass classifications.

The rest of the paper is structured as follows: In the second part of the study, participants, cognitive tasks, data collection and preprocessing and classification steps are mentioned under the title of materials and methods. In the third part, the findings are given. In the fourth chapter, results and evaluation sections are given.

## Objectives

The primary purpose of Brain-Computer Interface (BCI) studies is to establish communication between disabled individuals, other individuals, and machines with brain signals. Interpreting and classifying the brain's response during different cognitive tasks will contribute to brain-computer interface studies. Therefore, in this study, five cognitive tasks were classified from EEG signals.

## Materials and Methods

The general block diagram of the study is given in Figure 1. The study consists of three basic steps. These steps are collecting data, preprocessing data and classifying them in deep learning algorithms. When the experimental paradigm was examined, the EEGs of the volunteers were recorded at the time of rest and during the tests during the data collection phase. The resting moment lasts 45 seconds and the volunteers are given 15-second commands such as eye-open, eye-close and eye-open. In the cognitive activities phase, 5 tasks were applied to the volunteers in order and a 1-minute break was given between each task. The execution times of the tasks are average values. Depending on the success of the volunteers, the task completion time of each volunteer may vary. The implemented tasks are explained in section 2.2.

### Participants

30 healthy adult volunteers participated in the study. Volunteers are right-handed individuals who do not have any neurological or psychological diseases. Ethics committee approval of the study was obtained from Kütahya Health Sciences University

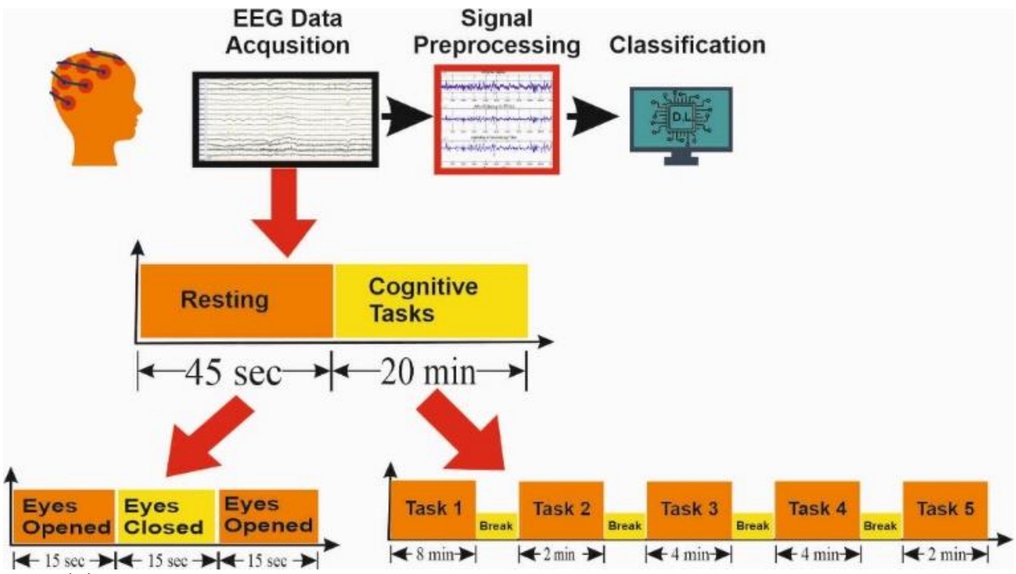


**Figure 1.** Experimental design.

Clinical Research Ethics Committee and each volunteer signed an informed consent form. The study was conducted in Turkey and all participants are Turkish.

### Cognitive Tasks

Within the scope of this study, five different neuropsychological tests were applied while collecting the EEG signals of the volunteers. Tests administered: Öktem verbal memory processes test, WMS-R visual memory subtest, Digit span test, Corsi block test and Stroop test.

#### Öktem verbal memory processes test [10], [11]

It is a word test. Volunteers are asked to keep the words in mind. The verbal memory and learning abilities of the volunteers are measured.

#### WMS-R visual memory subtest [12]

This test is a visual memory test. Volunteers are asked to draw the pictures shown after they are closed. The visual memory and learning abilities of volunteers are measured.

#### Digit Span test [13], [14]

This test is a verbal attention test and consists of two stages. It consists of the steps of repeating the number sequences read to the volunteers in the same way after the researcher and keeping them in mind and saying them from the end to the beginning.

Repeating numbers forward shows the efficiency and capacity of attention and concentration. Saying numbers backward is an execution task that depends on working memory.

#### Corsi Block test [15]

This test is a visual attention test and consists of two stages. It consists of touching the squares shown to the volunteers after the researcher, and touching them from the end to the beginning, keeping them in mind. The number of blocks touched for forward recall is evaluated as the person's visuospatial memory space. Tapping blocks backwards measures working memory capacity.

#### Stroop test [16], [17]

The Stroop test is a test that measures the brain's ability to direct attention, conceptual flexibility, and the processing speed of the mind. During this test, the volunteers are asked to read five cards or to say the colors of what is written on the cards.

### Data Acquisition and Preprocessing

EEG signals were recorded with a 16-channel Nihon Kohden EEG device using an international 10-20 electrode system according to bipolar electrodes. In Figure 2, there is a visual showing the bipolar electrode system. The sampling frequency of the EEG recording device is 500 Hz.

Electrode impedances were kept below 10kΩ. Artifacts of eye movements were removed by the ICA method [18],[19]. With the help of a 50Hz notch filter, mains noise is cleaned [20]. Data augmentation was applied to EEG data with amplifying all time data method and EEG data was tripled [21], [22]. Before starting the EEG recording, the volunteers were rested for a certain period of time, and the EEG recording was checked by giving eye-open and eye-close commands. All preprocessing steps are done using Python programming language and pandas, NumPy and mne packages. [23].

### Classification

The five tests performed by the volunteers were classified using two different deep learning algorithms. These deep learning algorithms are CNN and LSTM. In Figure 3, the learning model of the CNN network is given. The EEG signals passed

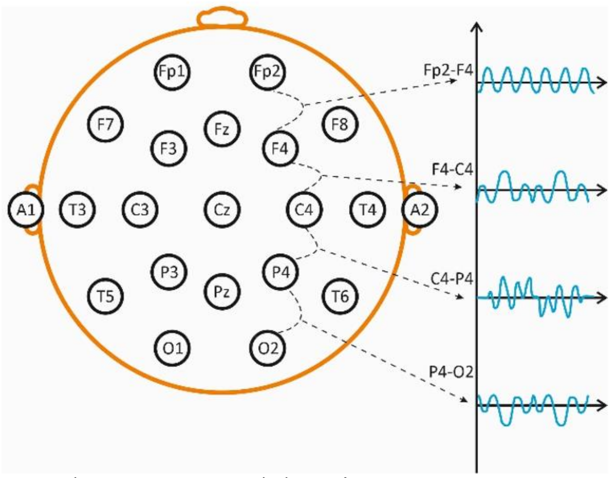


**Figure 2.** Bipolar electrode arrangement with 10-20 International electrode arrangement.

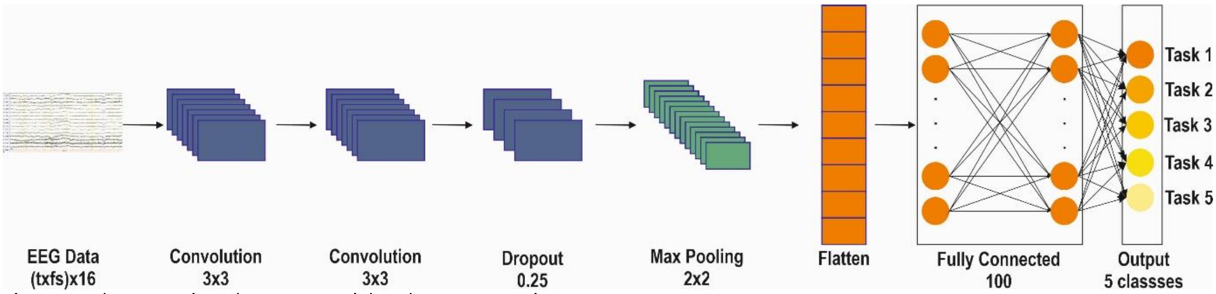


**Figure 3.** The CNN deep learning model architecture used.

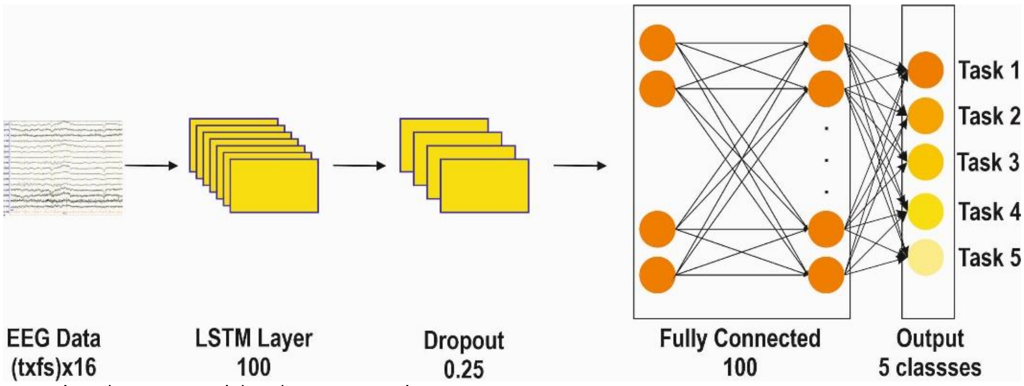


**Figure 4.** The LSTM deep learning model architecture used.

through the signal processing pre-steps are given as input to the network. Input signals (txfs) are matrices of x16 size. The t parameter, which determines the input matrix size, shows the duration of the EEG signals, the fs sampling frequency, and the 16-channel number. In the input of the CNN network model, two convolutional layers with 64 filters and a 3x3 window size are used. Then, one dropout layer of 25% and a maximum pooling layer of 2x2 size are used. At the output, it consists of a flatten layer, a fully connected layer and an output layer, respectively.
In Figure 4, the deep learning model of the LSTM network is given. The EEG signals passed through the signal processing pre-steps are given as input to the network. The LSTM layer is used as the first layer of the network. Then, overfitting was tried to be prevented with a 25% dropout layer. In the output, a fully connected dense layer with rectified linear activation function and the output layer with SoftMax activation function was used.

## Results

Five different cognitive tasks of the volunteers were classified by two different deep learning algorithms. 60% of the preprocessed EEG data was used for training in CNN and LSTM network models, 20% for validation and the other 20% for testing. The graph showing the training and validation results of the trained CNN and LSTM networks is given in Figure 5 and Figure 6. Figure 5 shows the training-validation process of the CNN network. The blue and orange lines in these figures illustrate the

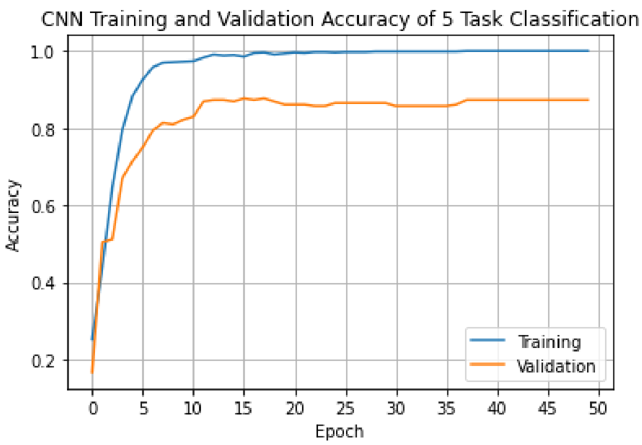


**Figure 5.** CNN model accuracy on training and validation sets of data (5 classes)

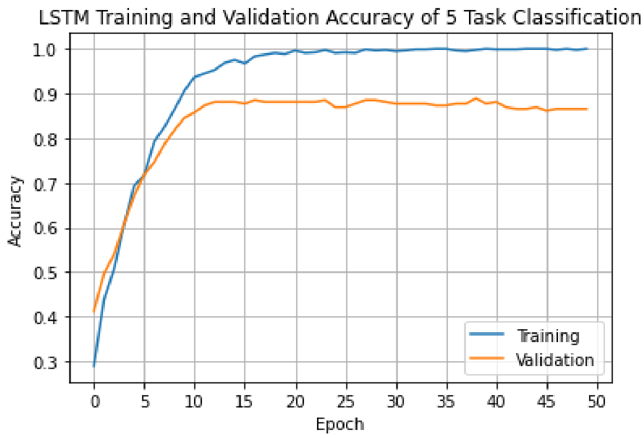


**Figure 6.** LSTM model accuracy on training and validation sets of data (5 classes)

training and validation process respectively. The proposed scheme has the advantage of accelerating convergence and reducing overfitting, as evidenced by the accuracy history in Fig. 5. As shown in these figures, our proposed CNN model reached its maximum performance very quickly, and the training-validation accuracy is stable.

Figure 6 shows the training-validation process of the LSTM network. The blue and orange lines in these figures illustrate the training and validation process respectively. Our proposed LSTM model reached its maximum performance very quickly, and the training-validation accuracy is stable.

The values showing the test metrics of CNN and LSTM networks are given in Table 1.When Table 1 is examined, the test accuracy for CNN is 88.53%. The Precision value is 88%, the Recall value is 87%, and the F1-Score value is 87%. The test accuracy for the LSTM network is 89.80%. The Precision value is 90%, the Recall value is 89%, and the F1-Score value is 89%. When we look at the other studies in the literature, there are studies in which multitasking based on physical performance is classified, cognitive tasks are not classified too much, or rather two cognitive tasks are classified.

**Table 1.** Accuracy metrics for the test data set.

| | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| CNN | 88.53 | 0.88 | 0.87 | 0.87 |
| LSTM | 89.80 | 0.90 | 0.89 | 0.89 |

In a study [24], in which hand, foot and tongue movements of volunteers were classified, right hand, left hand, feet and tongue movements were classified by deep learning algorithms. The average highest accuracy of 73.4% was found. In another study in which mental tasks were classified [25],two mental tasks were classified with SVM with an accuracy of 90.39%, while four mental tasks were classified with an accuracy of 80.09%.

In another recent study [26], five tasks of a test were classified with the Classification and regression tree (CART) and found an accuracy of 96.88%. In this study, five different cognitive tasks were classified by deep learning algorithms. When the confusion matrices of deep learning algorithms are examined, the confusion matrices of CNN, the test data set is in Figure 7. The real categories (columns) and predicted categories (rows) of the classification results can be read directly.

| Output Class | | Target Class: Corsi | Digit | Stroop | Verbal | Visual | |
|---|---|---|---|---|---|---|---|
| | Corsi | 71 | 7 | 0 | 0 | 2 | 88.75% / 11.25% |
| | Digit | 7 | 63 | 0 | 0 | 0 | 90.00% / 10.00% |
| | Stroop | 5 | 2 | 33 | 2 | 0 | 78.57% / 21.43% |
| | Verbal | 5 | 0 | 0 | 22 | 1 | 78.57% / 21.43% |
| | Visual | 4 | 0 | 0 | 1 | 27 | 84.38% / 15.62% |
| | Accuracy | 77.17% | 87.5% | 100% | 88% | 90% | 88.53% |
| | Error | 22.83% | 12.5% | 0% | 12% | 10% | 11.47% |

**Figure 7.** CNN confusion matrix for five classes.

| Output Class | | | | | | |
|---|---|---|---|---|---|---|
| Corsi | 71 | 3 | 5 | 1 | 0 | 88.75% 11.25% |
| Digit | 6 | 63 | 0 | 1 | 0 | 90.00% 10.00% |
| Stroop | 6 | 1 | 35 | 0 | 0 | 83.33% 16.67% |
| Verbal | 5 | 2 | 0 | 21 | 0 | 75.00% 25.00% |
| Visual | 2 | 0 | 0 | 0 | 30 | 93.75% 6.25% |
| Accuracy | 78.88% | 91.30% | 87.5% | 91.30% | 100% | 89.80% |
| Error | 21.12% | 8.7% | 12.5% | 8.7% | 0% | 10.20% |
| | Corsi | Digit | Stroop | Verbal | Visual | |
| | Target Class | | | | | |

**Figure 8.** LSTM confusion matrix

The confusion matrices of LSTM the test data set are in Figure 8. The real categories (columns) and predicted categories (rows) of the classification results can be read directly.

## Discussion

When we review other studies in the literature, there are studies in which multitasking based on physical performance is classified, cognitive tasks are not classified too much, or rather two cognitive tasks are classified. In a study [24], in which hand, foot, and tongue movements of volunteers were classified, right hand, left hand, feet, and tongue movements were classified by deep learning algorithms. The average highest accuracy of 73.4% was found. In another study in which mental tasks were classified [25], two mental tasks were classified with SVM with an accuracy of 90.39%, while four mental tasks were classified with an accuracy of 80.09%. In another recent study [26], five tasks of a test were classified with the classification and regression tree (CART) and found an accuracy of 96.88%. In this study, five different cognitive tasks were classified by deep learning algorithms. When the confusion matrices of deep learning algorithms are examined, the confusion matrices of CNN, the test data set is in Figure 7. The real categories (columns) and predicted categories (rows) of the classification results can be read directly.

## Conclusions

Within the scope of the study, 5 cognitive tests (Öktem verbal memory processes test, WMS-R Visual Memory subtest, Digit span test, Corsi Blok test and Stroop test) measuring verbal memory, visual memory, attention, concentration, working memory and reaction time abilities of 30 healthy volunteers) recorded EEG signals were classified using deep learning methods. In the classification results, 5 different tests within the scope of the study were classified with 88.53% accuracy with the CNN algorithm, while with the LSTM deep learning algorithm, they could be distinguished with an accuracy of 89.80%. Precision, recall and f1 score were used as evaluation metrics. The study was conducted on healthy volunteers and can be extended by repeating it on individuals with cognitive or physical impairments. If repeated with high accuracy on sick individuals, it will be a study that will form the basis for BCI studies. In future studies, the number of volunteers and the number of tests can be increased, the tests can be collected more quickly in the computer environment.

### Compliance with ethical guidelines

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants involved in the study..


### Acknowledgments

This study was supported by the Scientific Research Projects of Kütahya Dumlupınar University within the scope of the project numbered 2020/26. Also, it was conducted by researchers at Kütahya Dumlupınar University Neurotechnology Education Application and Research Center (NÖTEM).


### Authors' contributions

Evin Şahin Sadık: Methodology, Software, Investigation, Formal analysis, Resources, Data curation, Visualization, Writing.
Hamdi Melih Saraoğlu: Conceptualization, Methodology, Supervision, Project administration, Funding acquistion, Writing - Review & Editing.
Sibel Canbaz Kabay: Conceptualization, Methodology, Supervision, Writing - Review & Editing.
Cahit Keskinkılıç: Methodology, Formal analysis, Investigation, Writing - Review & Editing.


### Funding/Support

This study was supported by the Scientific Research Projects of Kütahya Dumlupınar University within the scope of the project numbered 2020/26.


### Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References


1. B. Blankertz *et al.*, "The BCI competition 2003: Progress and Blankertz B, Muller KR, Curio G, Vaughan TM, Schalk G,

Wolpaw JR, et al. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Transactions on Biomedical Engineering. 2004; 51(6):1044-51. [DOI:10.1109/TBME.2004.826692] [PMID]
2. Gupta A, Agrawal RK, Kaur B. Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods. Soft Computing. 2015; 19:2799-812. [DOI:10.1007/s00500-014-1443-1]
3. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, et al. Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering. 2000; 8(2):164-73. [DOI:10.1109/TRE.2000.847807] [PMID]
4. Suchetha M, Madhumitha R, Sruthi R. Sequential Convolutional Neural Networks for classification of cognitive tasks from EEG signals. Applied Soft Computing. 2021; 111:107664. [DOI:10.1016/j.asoc.2021.107664]
5. Li W, Ming D, Xu R, Ding H, Qi H, Wan B. Research on visual attention classification based on EEG entropy parameters. InWorld Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China; 2013.
6. Amin HU, Mumtaz W, Subhani AR, Saad MN, Malik AS. Classification of EEG signals based on pattern recognition approach. Frontiers in computational neuroscience. 2017; 11:103. [DOI:10.3389/fncom.2017.00103] [PMID]
7. Fatimah B, Javali A, Ansar H, Harshitha BG, Kumar H. Mental arithmetic task classification using fourier decomposition method. In2020 International Conference on Communication and Signal Processing (ICCSP); 2020.
8. Guevara MA, Paniagua EI, González MH, Carrillo IK, Sepúlveda ML, Orozco JC, et al. EEG activity during the spatial span task in young men: Differences between short-term and working memory. Brain Research. 2018; 1683:86-94. [DOI:10.1016/j.brainres.2018.02.004] [PMID]
9. Kutlu Y, Yayık A, Yildirim E, Yildirim S. LU triangularization extreme learning machine in EEG cognitive task classification. Neural Computing and Applications. 2019; 31:1117-26. [DOI:10.1007/s00521-017-3142-1]
10. O. Oktem, “Verbal Memory Processes Test (SBST),” Neuro-psychiatry Archive, vol. 29, pp. 196–206, 1992.
11. O. O. Tanor, “Oktem verbal memory processes test(Oktem-SBST) handbook.” Turkish Psychological Association, 2011.
12. Wechsler D. Wechsler memory scale-revised. San Antonio. Harcourt Brace Jovanovich; 1987.
13. S. Karakas and A. Yalin, “The standardization study of the audio-visual number sequences test form B on the 13-54 age group,” Turkish Psycho. Jour., vol. 10, no. 34, pp. 20–31, 1995.
14. U. M. Osmanlioglu and M. Ozguzel, “Examination of patients in various diagnostic groups with memory impairment with the Wechsler Memory Scale,” XXI. Psychiatrist. and Neurological Science. Congress Book, Adana, Cukurova University Yayını, 1985.
15. Corsi PM. Human memory and the medial temporal region of the brain. [PhD Thesis]. McGill University; 1972.
16. Jensen AR, Rohwer Jr WD. The Stroop color-word test: a review. Acta psychologica. 1966; 25(1):36-93. [DOI:10.1016/0001-6918(66)90004-7] [PMID]
17. S. Karakas, E. Erdogan, A. S. Soysal, T. Ulusoy, I. Ulusoy, and S. Alkan, “Stroop Test TBAG Form: Standardization Studies for Turkish Culture ,” Turkish Psychiatrist Journal, vol. 2, pp. 75–88, 1999.
18. Vigário R, Sarela J, Jousmiki V, Hamalainen M, Oja E. Independent component approach to the analysis of EEG and MEG recordings. IEEE transactions on biomedical engineering. 2000; 47(5):589-93. [DOI:10.1109/10.841330] [PMID]
19. Escudero J, Hornero R, Abásolo D, Fernández A, López-Coronado M. Artifact removal in magnetoencephalogram background activity with independent component analysis. IEEE Transactions on Biomedical Engineering. 2007; 54(11):1965-73. [DOI:10.1109/TBME.2007.894968] [PMID]
20. Luck SJ. An introduction to the event-related potential technique. MIT press; 2014.
21. Lashgari E, Liang D, Maoz U. Data augmentation for deep-learning-based electroencephalography. Journal of Neuro-science Methods. 2020; 346:108885. [DOI:10.1016/j.jneumeth.2020.108885] [PMID]
22. Sakai A, Minoda Y, Morikawa K. Data augmentation methods for machine-learning-based classification of bio-signals. In2017 10th Biomedical Engineering International Conference (BMEiCON); 2017.
23. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience. 2013; 7:267. [DOI:10.3389/fnins.2013.00267] [PMID] [PMCID]
24. Song Y, Wang D, Yue K, Zheng N, Shen ZJ. EEG-based motor imagery classification with deep multitask learning. In2019 International Joint Conference on Neural Networks (IJCNN); 2019.
25. Lim WL, Sourina O, Liu Y, Wang L. EEG-based mental workload recognition related to multitasking. In 10th International Conference on Information, Communications and Signal Processing (ICICS); 2015.
26. Sahin Sadik E, Saraoglu HM, Canbaz Kabay S, Tosun M, Keskinkilic C, Akdag G. Investigation of the effect of rosemary odor on mental workload using EEG: an artificial intelligence approach. Signal, Image and Video Processing. 2022; 16(2):497-504. [DOI:10.1007/s11760-021-01992-5]