COMPUTER RESEARCH AND MODELING
2023 VOL. 15 NO. 3 P. 675–690
DOI: 10.20537/2076-7633-2023-15-3-675-690

КиМ

ANALYSIS AND MODELING OF COMPLEX LIVING SYSTEMS

UDC: 519.6, 519.7, 004.5

Optimization of the brain command dictionary based on the statistical proximity criterion in silent speech recognition task

A. Bernadotte[1,2,3,a], A. D. Mazurin[2,b]

[1]National University of Science and Technology MISIS,
4 Leninskiy pr., Moscow, 119049, Russia
[2]Faculty of Mechanics and Mathematics, Moscow State University,
GSP-1, Leninskie Gory, Moscow, 119991, Russia
[3]LLC Neurosputnik,
96 pr. Vernadskogo, Moscow, 119571, Russia

E-mail: [a] bernadotte.alexandra@intsys.msu.ru, [b] mazurin1567@gmail.com

Received 06.01.2023, after completion — 10.04.2023.
Accepted for publication 10.05.2023.

In our research, we focus on the problem of classification for silent speech recognition to develop a brain–computer interface (BCI) based on electroencephalographic (EEG) data, which will be capable of assisting people with mental and physical disabilities and expanding human capabilities in everyday life. Our previous research has shown that the silent pronouncing of some words results in almost identical distributions of electroencephalographic signal data. Such a phenomenon has a suppressive impact on the quality of neural network model behavior. This paper proposes a data processing technique that distinguishes between statistically remote and inseparable classes in the dataset. Applying the proposed approach helps us reach the goal of maximizing the semantic load of the dictionary used in BCI.

Furthermore, we propose the existence of a statistical predictive criterion for the accuracy of binary classification of the words in a dictionary. Such a criterion aims to estimate the lower and the upper bounds of classifiers' behavior only by measuring quantitative statistical properties of the data (in particular, using the Kolmogorov – Smirnov method). We show that higher levels of classification accuracy can be achieved by means of applying the proposed predictive criterion, making it possible to form an optimized dictionary in terms of semantic load for the EEG-based BCIs. Furthermore, using such a dictionary as a training dataset for classification problems grants the statistical remoteness of the classes by taking into account the semantic and phonetic properties of the corresponding words and improves the classification behavior of silent speech recognition models.

Keywords: brain–computer interface, EEG, silent speech classification, graph dictionary selection algorithm, BCI, deep learning optimization, silent speech recognition, statistical proximity criterion

Citation: *Computer Research and Modeling*, 2023, vol. 15, no. 3, pp. 675–690.

КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2023 Т. 15 № 3 С. 675–690
DOI: 10.20537/2076-7633-2023-15-3-675-690

Ки&М

АНАЛИЗ И МОДЕЛИРОВАНИЕ СЛОЖНЫХ ЖИВЫХ СИСТЕМ

УДК: 519.6, 519.7, 004.5

Оптимизация словаря команд на основе статистического критерия близости в задаче распознавания невербальной речи

А. К. Бернадотт[1,2,3,a], А. Д. Мазурин[2,b]

[1]Кафедра инженерной кибернетики, Национальный исследовательский технологический университет «МИСиС»,
Россия, 119049, Москва, Ленинский прю, д. 4, стр. 1
[2]Механико-математический факультет МГУ,
Россия, 119991, Москва, Ломоносовский пр., д. 1, ГЗ
[3]ООО Нейроспутник,
Россия, 119571, Москва, пр. Вернадского, д. 96

E-mail: [a] bernadotte.alexandra@intsys.msu.ru, [b] mazurin1567@gmail.com

Получено 06.01.2023, после доработки — 10.04.2023.
Принято к публикации 10.05.2023.

В исследовании мы сосредоточились на задаче классификации невербальной речи для разработки интерфейса «мозг–компьютер» (ИМК) на основе электроэнцефалографии (ЭЭГ), который будет способен помочь людям с ограниченными возможностями и расширить возможности человека в повседневной жизни. Ранее наши исследования показали, что беззвучная речь для некоторых слов приводит к почти идентичным распределениям ЭЭГ-данных. Это явление негативно влияет на точность классификации нейросетевой модели. В этой статье предлагается метод обработки данных, который различает статистически удаленные и неразделимые классы данных. Применение предложенного подхода позволяет достичь цели максимального увеличения смысловой нагрузки словаря, используемого в ИМК.

Кроме того, мы предлагаем статистический прогностический критерий точности бинарной классификации слов в словаре. Такой критерий направлен на оценку нижней и верхней границ поведения классификаторов только путем измерения количественных статистических свойств данных (в частности, с использованием метода Колмогорова – Смирнова). Показано, что более высокие уровни точности классификации могут быть достигнуты за счет применения предложенного прогностического критерия, позволяющего сформировать оптимизированный словарь с точки зрения семантической нагрузки для ИМК на основе ЭЭГ. Кроме того, использование такого обучающего набора данных для задач классификации по словарю обеспечивает статистическую удаленность классов за счет учета семантических и фонетических свойств соответствующих слов и улучшает поведение классификации моделей распознавания беззвучной речи.

Ключевые слова: интерфейс «мозг–компьютер», ЭЭГ, классификация невербальной речи, графовый алгоритм выбора словаря, ИМК, оптимизация глубокого обучения, распознавание невербальной речи, статистический критерий близости

Introduction

In our work, we are developing a brain-computer interface (BCI) based on electroencephalography (EEG), which focuses on recognizing movement mental commands. This interface is especially important for people with damage or underdevelopment of the motor cortex. Neurodegenerative diseases and natural ageing are often associated with chronic inflammation in the brain tissue, loss of brain plasticity, and loss of brain function—precisely in the specific brain areas responsible for motor function [Karpenko et al., 2018; Bernadotte, Mikhelson, Spivak, 2016; Bernadotte et al., 2014; Vasilishina et al., 2019; Aarsland, Bernadotte, 2015]. Moreover, many neurological diseases, such as amyotrophic lateral sclerosis, Parkinson's disease, multiple sclerosis, and central nervous system injuries, are known to be accompanied or followed by communication and movement dysfunction. Recent research shows that the long-term pathological effects of COVID-19 may cause such problems. In particular, our colleagues [Zubov, Isaeva, Bernadotte, 2021] showed that signal patterns detected by electroencephalography could indicate the peculiarities of brain functioning after suffering COVID-19. Therefore, there is a constantly growing need for devices such as BCIs that would help in rehabilitation or adaptation to the loss of communication or movement capabilities of patients with neurological diseases.

Since non-invasive BCIs are most often designed based on EEG, the modality imposes serious restrictions on the areas of use of the BCIs and the applied processing algorithms. The EEG signal contains a large proportion of noise; the signal is projective and correlated across channels.

Bearing this problem in mind, we decided to develop a BCI that recognizes silent speech, that is, patterns of brain activity in the temporal lobe of the cortex, or rather, in Broca's area (the anterior speech cortex). In our study, the silent speech was a set of commands (dictionary) given by the inner voice. These commands should not be numerous to control the manipulator, but the BCI should recognize the dictionary well.

When developing BCIs to help people with disabilities, a mental signal of silent speech or movement is often used. Recognition of mental commands (silent speech) is one of the classic methods for designing interfaces. However, until recently, the accuracy of command recognition was very low.

A vital breakthrough in BCIs development was using neural networks for silent speech classification. For a long time, all scientific thought in this area was moving towards the proposal of neural network architectures. However, after a certain accuracy had been reached, it was impossible to move further. For a long time, the accuracy of the binary classification of phonemes or silent words did not exceed 86 % [Pandey, Wang, 2019; Kuchaiev, Boris, 2017; Bromley et al., 1994; DaSalla et al., 2009; Brigham, Kumar, 2010; Min et al., 2016; Huang, Zhu, Siew, 2004; Balaji et al., 2017; Nguyen, Karavas, Artemiadis, 2018; Cooney, Folli, Coyle, 2018; Panachakel, Ramakrishnan, Ananthapadmanabha, 2019; Pramit, Muhammad, Sidney, 2019; Ossadtchi, Lebedev, 2020; Lebedev, 2019]. A detailed analysis can be found in our article [Vorontsova et al., 2021].

Our scientific group contributed to the development of BCIs and proposed methods that significantly increased the mental commands classification accuracy [Bernadotte, 2022a; Bernadotte, 2022b; Vorontsova et al., 2021]. In addition, we looked at sociocultural mental patterns and the consequences of COVID-19, and most importantly, we applied a graph algorithm to increase the accuracy dramatically.

In our previous work [Mazurin, Bernadotte, 2021], we showed that people belonging to several common social categories share similar electroencephalographic signal patterns, allowing us to divide the dataset into corresponding groups and train neural network classification models separately for each of these groups. However, even such a procedure does not always provide an increase in classification accuracy. Here our reasoning went further, and we decided to inspect the statistical properties of the collected EEG data. In particular, it was found that some commands pronounced by the inner

voice (silent speech) are poorly distinguishable from each other and demonstrate a low accuracy of classification by a neural network of BCI. We hypothesized that this silent speech classification problem is due to the semantic and phonetic proximity of the words. It was shown that the silent speech of some words leads to almost identical distributions of electroencephalographic signal data [Vorontsova et al., 2021].

Earlier, we showed that this silent speech classification problem could be overcome by selecting a dictionary of well-recognized and accurately classified words. Previously, our group presented a graph dictionary selection algorithm and the applicability of this algorithm to our data [Bernadotte, 2022a; Bernadotte, 2022b]. The situation with selecting such a dictionary is justified by the purpose of the work, which is to create a device that recognizes mental commands corresponding to different motion directions.

The selection of a dictionary itself is a rather laborious task, and we had an idea to develop a certain criterion that would allow us to predict the behavior of a neural network and, at the same time, would be quite convenient in terms of time and resource. Thus, we were looking for a criterion in statistical methods that would be less expensive in calculating resources.

The new study presented in this paper is based on the analysis of classification behavior of neural network models trained on the set of EEG data recordings during several sessions corresponding to silent speech pronunciation of words “up”, “down”, “vira”, and “myna”. These words are semantically divided into two classes: {“up”, “vira”} and {“down”, “myna”}. To avoid time- and GPU resource-consuming dictionary selection, it was interesting for us to consider the existence of a predictive criterion by which we could predict the results of the binary classification of a neural network.

Thinking about the predictive criterion for neural network classification accuracy, we formulated the following hypothesis:

Hypothesis 1. *The Kolmogorov – Smirnov method on EEG data of silent speech with reduced dimension can statistically predict a neural network classifier’s behavior* (*accuracy*) *on the same EEG data of a higher dimension.*

Confirming this hypothesis and gaining new knowledge about the patterns of brain activity aims at improving our classification results, not only by trying to tune various neural network architectures but also by intelligently tuning the dictionary with particular attention to the statistical proximity of the classes in the dictionary.

Moreover, proving the hypothesis directly leads to the fact that there exists a predictive criterion for the classification accuracy of the models based on the overall statistical separability of the classes in the dataset. In particular, we show that the sum of all p-values computed pairwise for every two classes in the train dataset can be used to make an estimation of the accuracy rate levels of a given network.

We were looking for our criterion in the field of statistical methods that were less expensive in terms of calculating resources.

Methods

Dataset

All subjects had reached adulthood, were healthy, and voluntarily signed a consent to the study. The subjects had the right to withdraw from the study at any time without explanation. The subjects provided their data, which included: gender, age, education, and occupation. The exclusionary criteria for the study were a history of head trauma, alcohol or other intoxication, and epilepsy.

The dataset we used for our study consisted of 32-channel EEG signal recordings completed at 250 Hz during several sessions of silent and vocalized speech of 105 subjects. The dry plastic

electrodes (Datwyler's SoftPulseTM Medium, brush type electrode) were placed according to the traditional 10–20 scheme. The "Afz"-channel was used as a reference electrode. A word presentation signal was also captured via a light sensor and included in data files as a mark.

Data processing

Data preprocessing methods included: 1) eye noise filtering procedure consisting of morphological selection of eye sensitive channels, Independent Component Analysis (ICA), and further detection and removal of the eye noise component using Fast Fourier Transform (FFT), Savitskiy – Galey filtering and Inverse FFT; 2) filtering out the tensors with a high noise level by applying two filters by the sum of the moduli of the signal amplitudes; 3) downsampling the sample rate by using index masks on the original EEG data; 4) separating the electrodes into left and right hemispheres; 5) reorganizing the elements of the data tensors and combining the operations described above in order to form a two-dimensional matrix from every tensor in the dataset [Vorontsova et al., 2021].

By applying the procedures described above, we were able to transform any raw signal tensor of size 32×1024 (32 stands for the number of electrodes (channels) used in the EEG recording procedure; 1024 is the number of time stamps corresponding to signal discretization procedure) to a square 2D matrix of size 256×256.

Predictive criterion

Before using the criterion, it is strongly recommended to use principal component analysis (PCA). After PCA, we looked at distributions of mental words as distributions in kD-space. We used 3D-space. The Kolmogorov – Smirnov test, independent of the nature of the distribution, was applied to the set of distributions in 3D-space by computing the set of pairwise p-values on train, validation, and test sets. Each component (each dimension) had its own p-value, the minimum p-value was taken from this set. The existing pairs of synonyms were organized in ascending order of p-values.

The predictive criterion was based on statistics of dimension reduced data: on the set of all possible p-values of the Kolmogorov – Smirnov test on PCA dimension reduced distributions from validation and train sets, the p-value and the network binary classification accuracy are inversely correlated. This absolute accuracy (average or median) depends on the network architecture and has a variance depending on the network architecture.

Neural Network Classifiers

Neural network classification models which were tested for multi-class classification of various sets of word commands included Image Transformer and ResNext 18 deep architectures adapted for the preprocessed dataset consisting of square 256×256 tensors. We further briefly describe the above-mentioned deep learning architectures.

ResNext-18 Network Classifier

The main feature that makes ResNext model differ from conventional residual nets is the introduction of the cardinality hyperparameter [Xie et al., 2016], which separates the channels of the input tensors into several groups, with each one being operated by its own convolutional kernel. This architecture adopts both the strategy of repeating layers, which is a common property of VGGs/ResNets, and the split-transform-merge strategy first applied in Inception model architectures. ResNext models are mainly constructed from building blocks, each performing a set of convolutional transformations on a low-dimensional embedding and aggregating their outputs by summation [Xie et al., 2016]. Model configurations differ in complexity and the number of building blocks in the architecture. We chose the ResNext-18 model, which consists of 4 modules, each containing two convolutional blocks (see Fig. 1 for a detailed architecture scheme).

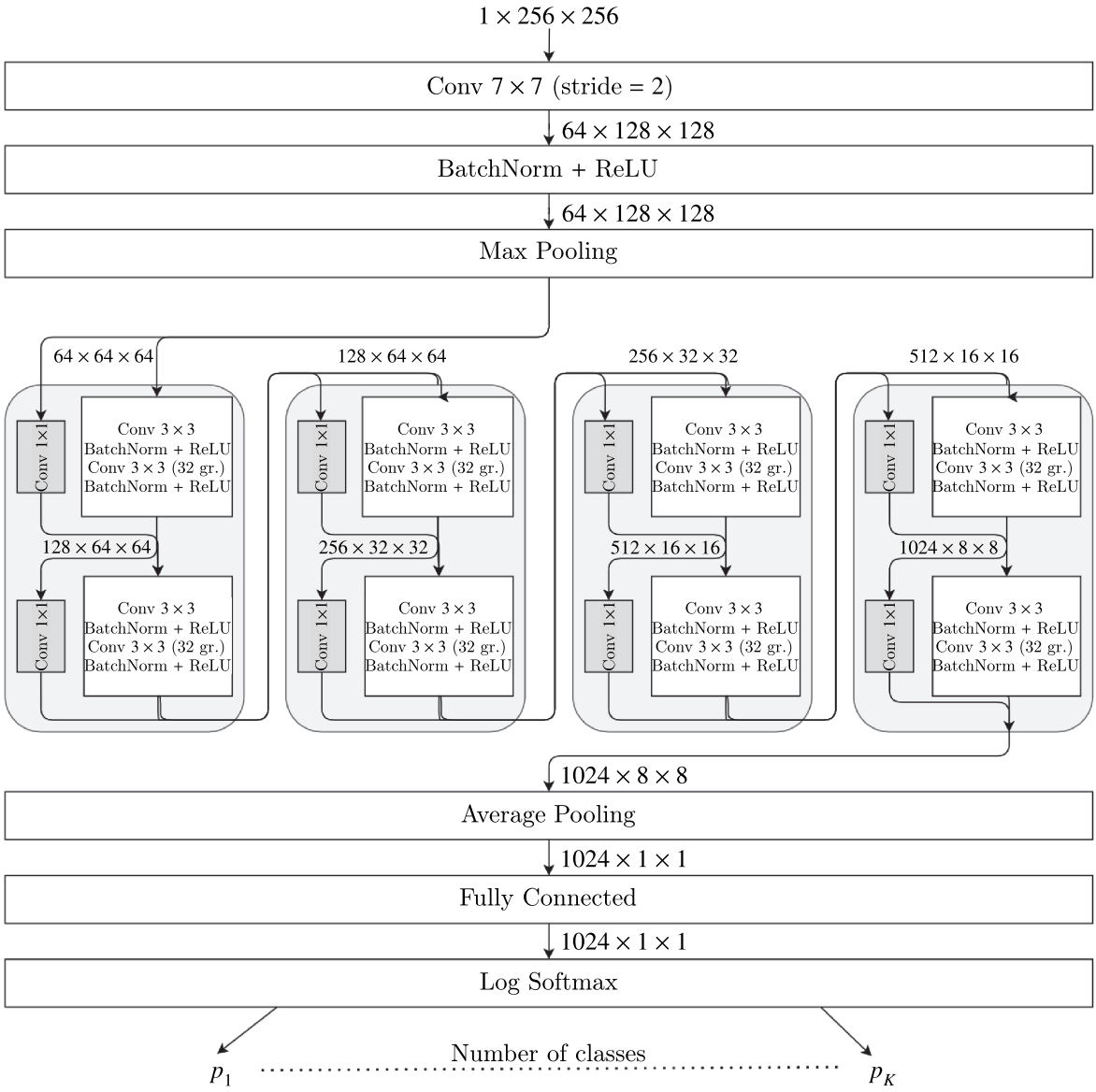

Figure 1. ResNext-18 model architecture

Vision Transformer Network Classifier

We decided to use the Vision Transformer neural network classifier based on colleagues' paper [Dosovitskiy et al., 2020] as well as convolution-based architectures. This version of transformer architecture is specifically designed for the task of image classification. Vision Transformer contains a solid number of self-attention blocks, the main purpose of which is a computation of point-wise scalar products between the values of all square (16×16) fragments of the preprocessing 256×256 matrix and vectors consisting of trainable weights. Such a solution may allow the network to find hidden patterns in the information provided by the whole input matrix, not only in its neighboring fragments, which grants it a benefit when compared to convolutional neural networks (CNNs). Moreover, the self-attention blocks of Vision Transformer net are connected successively, forming a chain which makes the process of finding deeper patterns easier.

The architecture of Vision Transformer (see Figure 2) is aimed at dealing with square images by means of initial cropping into a number of square patches of smaller size. These patches serve as inputs to the trainable part of the network, which mainly consists of self-attention blocks responsible

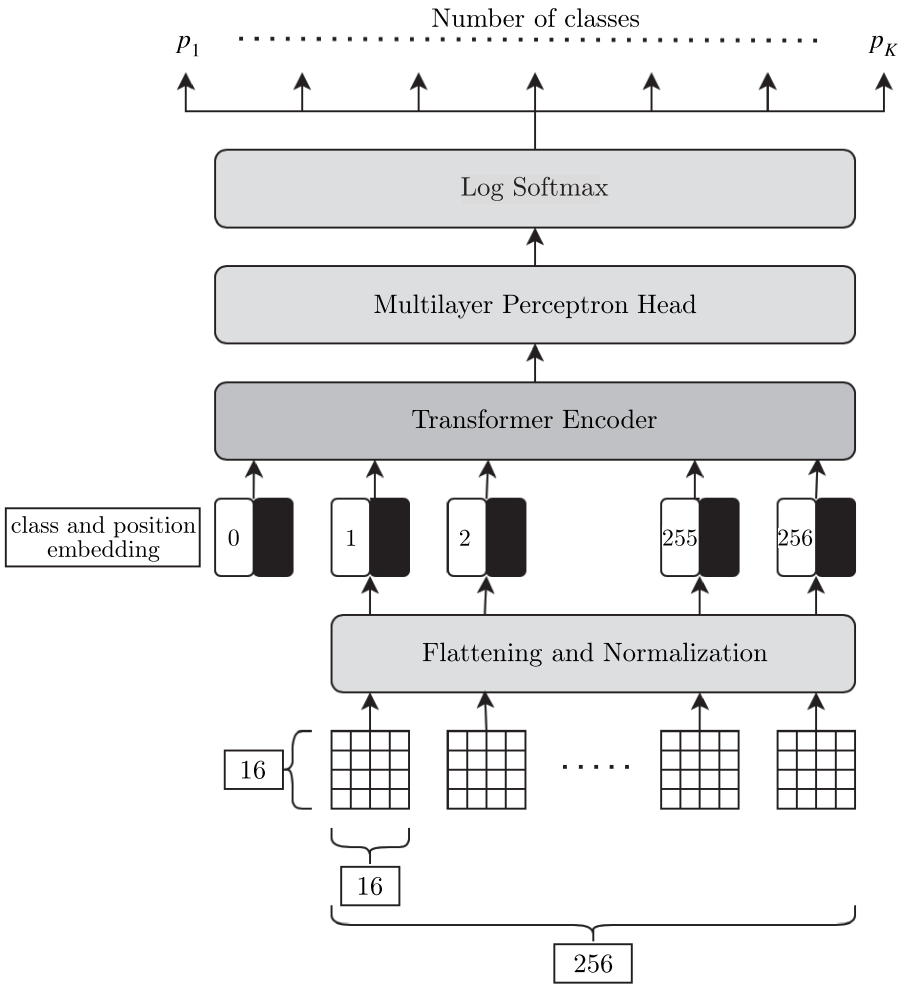

Figure 2. The architecture of Vision Transformer

for finding connections between them and deciding which of them are most important for completing the process of classification.

First, the inputs of the net are cropped into 256 square patches of size 16×16, which are further flattened, normalized, and enter the embedding layer immediately after that. At this stage, a positional token with learnable parameters is added to each of the inputs. Next, the embedded tensor patches serve as inputs for the main part of the network — successively connected one after another N Transformer Encoder blocks (see Figure 3). Inside each of these blocks there is a multi-head self-attention layer (taken exactly from [Dosovitskiy et al., 2020]) followed by a multilayer perceptron block, which consists of two fully connected layers forming a narrow bottleneck with dropout and GELU activation. We use $N = 6$ transformer encoder blocks, $h = 8$ as the number of neurons in the hidden layer of the MLP block and $p = 6$ as the number of heads in the multi-head self-attention layer. The experiments showed that increasing the complexity of the model did not result in higher results for validation and test classification.

Neural Networks Training details

We trained all models, including ResNext-18 and Vision Transformer, using the Adam optimizing method with learning rate $\alpha = 0.001$ and parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and apply a high weight decay rate of $w = 0.01$, which proved to be useful to avoid overfitting for our dataset. We use negative

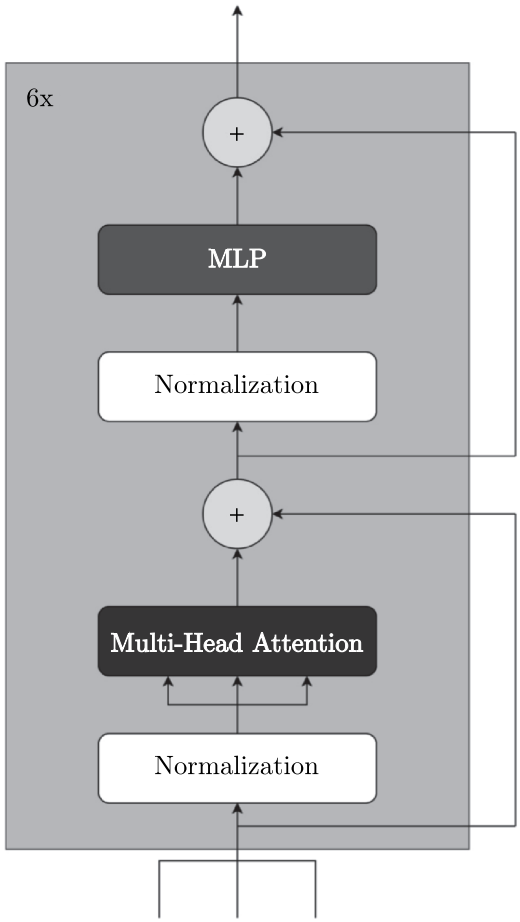

Figure 3. Transformer Encoder block structure

log likelihood as a loss function to optimize during neural network training. The size of each batch during the training stage is equal to 64. The data were divided into three sets: training (70 %), validation (20 %), and test (10 %). The training stage of the classifiers lasted 500 epochs. However, we found that both architectures tend to show satisfactory classification results after only 100 epochs of training. After each training epoch, we test the classification behavior of the models on the validation dataset obtained using the out-of-sample approach. Model parameters which result in the lowest validation loss value among all training epochs are saved and further used to estimate the classification score on the test part of the dataset. Moreover, we carried out each of the neural network training experiments with similar parameters and trained classes six times repeatedly with different random seeds, which are responsible for the weight initialization inside the layers of the network architectures. After completing each series of 6 experiments, we calculate the mean and median values of classification scores in order to obtain information about the medium behavior of the networks when trained on different sets of classes. We present the resulting accuracy scores of all experiments as well as their median values in the following paragraph.

Results

Kolmogorov – Smirnov test results

We applied the Kolmogorov – Smirnov test for every pair of classes in the initial dictionary (“up”, “down”, “vira”, “myna”), presented by their distributions in 3D-space. After the set of *p*-values

is computed for each of 3 components (dimensions) of the data, the minimum value is chosen from it to indicate the degree of statistical separability between two classes. The results of the Kolmogorov – Smirnov test, computed separately for training, validation, and test parts of the dataset, are presented in Tables 1–3, respectively.

Table 1. *P*-values obtained by the Kolmogorov – Smirnov test application to train part of the dataset

words (train)	"up"	"down"	"vira"	"myna"
"up'	—	$7.33 \times e^{-1}$	$1.72 \times e^{-174}$	$1.39 \times e^{-175}$
"down"	$7.33 \times e^{-1}$	—	$4.92 \times e^{-168}$	$4.92 \times e^{-168}$
"vira"	$1.72 \times e^{-174}$	$4.92 \times e^{-168}$	—	$4.99 \times e^{-1}$
"myna"	$1.39 \times e^{-175}$	$4.92 \times e^{-168}$	$4.99 \times e^{-1}$	—

Table 2. *P*-values obtained by the Kolmogorov – Smirnov test application to the validation part of the dataset

words (validation)	"up"	"down"	"vira"	"myna"
"up"	—	$3.34 \times e^{-1}$	$1.11 \times e^{-16}$	$1.52 \times e^{-52}$
"down"	$3.34 \times e^{-1}$	—	$1.11 \times e^{-16}$	$9.92 \times e^{-58}$
"vira"	$1.11 \times e^{-16}$	$1.11 \times e^{-16}$	—	$4.21 \times e^{-1}$
"myna"	$1.52 \times e^{-52}$	$9.92 \times e^{-58}$	$4.21 \times e^{-1}$	—

Table 3. *P*-values obtained by the Kolmogorov – Smirnov test application to test part of the dataset

words (test)	"up"	"down"	"vira"	"myna"
"up"	—	$1.61 \times e^{-1}$	$1.52 \times e^{-63}$	$1.52 \times e^{-63}$
"down"	$1.61 \times e^{-1}$	—	$1.52 \times e^{-63}$	$1.52 \times e^{-63}$
"vira"	$1.52 \times e^{-63}$	$1.52 \times e^{-63}$	—	$5.54 \times e^{-1}$
"myna"	$1.52 \times e^{-63}$	$1.52 \times e^{-63}$	$5.54 \times e^{-1}$	—

ResNext-18 binary classification results

We tested the work of the classifiers on various deep neural network architectures, including convolutional networks and transformers. Among convolutional architectures, we tried ResNet-18, ResNext-18, ResNext-50, ResNext-101, and ResNext-152. The highest score results were obtained by implementing the ResNext-18 neural network. The results of the application of the ResNext-18 network to our dataset are shown in Table 4. We created six copies of the network with various weight initialization and launched training procedures for each of them separately.

Vision Transformer binary classification results

In this paragraph, we present the results of experiments in the problem of binary classification carried out by Vision Transformer. As in the case of ResNext-18, we repeated the process of neural network training six times with different random seeds responsible for layer weight initialization. The results of Vision Transformer training are shown in Table 5.

We can see the same regularities in the set of the obtained test accuracy score as in the case of ResNext-18. In particular, pairs "up"/"down" and "vira"/"myna" have poor classification results (0.5, corresponding to the random guess), while all other pairs of classes yield significantly better behavior of

Table 4. ResNext-18 classification results

word pair	experiment number	total test accuracy	“up” accuracy	“down” accuracy	“vira” accuracy	“myna” accuracy
“up”/“down”	1	0.5	1.0	0.0	—	—
	2	0.5	0.0	1.0	—	—
	3	0.5	0.0	1.0	—	—
	4	0.5	0.0	1.0	—	—
	5	0.514	0.291	0.737	—	—
	6	0.517	0.034	1.0	—	—
“vira”/“myna”	1	0.5	—	—	1.0	0.0
	2	0.491	—	—	0.561	0.421
	3	0.5	—	—	0.595	0.404
	4	0.516	—	—	0.050	0.983
	5	0.469	—	—	0.595	0.342
	6	0.5	—	—	0.595	0.404
“up”/“myna”	1	0.719	0.811	—	—	0.629
	2	0.78187	0.942	—	—	0.623
	3	0.847	1.0	—	—	0.696
	4	0.694	0.382	—	—	1.0
	5	0.461	0.188	—	—	0.730
	6	0.662	0.577	—	—	0.747
“up”/“vira”	1	0.796	0.862	—	0.730	—
	2	0.504	0.88	—	0.134	—
	3	0.580	0.222	—	0.932	—
	4	0.804	0.88	—	0.730	—
	5	0.832	0.937	—	0.730	—
	6	0.694	0.811	—	0.578	—
“down”/“vira’	1	0.594	—	0.262	0.921	—
	2	0.597	—	0.188	1.0	—
	3	0.968	—	0.937	1.0	—
	4	0.923	—	0.914	0.932	—
	5	0.529	—	0.325	0.730	—
	6	0.779	—	0.954	0.606	—
“down”/“myna”	1	0.399	—	0.325	—	0.471
	2	0.767	—	0.977	—	0.561
	3	0.864	—	0.948	—	0.780
	4	0.810	—	0.942	—	0.679
	5	0.776	—	0.822	—	0.730
	6	0.787	—	0.982	—	0.595

the classifier. The overall level of accuracies is higher than for ResNext-18. Observe Table 6 to compare mean and median values of classification scores obtained by the neural networks described above.

Statistical predictive criterion

In this paragraph, we introduce the statistical criterion which establishes the connection between the p-values obtained from the Kolmogorov – Smirnov test and the accuracy scores of trained classifiers. First, we do it separately for the case of the two neural network classifiers described above. After that, we form the predictive criterion, which fits both architectures simultaneously, in order to prove the hypothesis stated in the introduction.

Predictive criterion for ResNext-18

Based on Tables 1–3 with p-values for different parts of the dataset and Table 6 containing median accuracy scores for ResNext-18, we are able to construct a criterion which reflects the

Table 5. Vision Transformer classification results

word pair	experiment number	total test accuracy	“up” accuracy	“down’ accuracy	“vira’ accuracy	“myna’ accuracy
“up”/“down”	1	0.491	0.811	0.171	—	—
	2	0.482	0.068	0.897	—	—
	3	0.491	0.817	0.165	—	—
	4	0.508	0.251	0.765	—	—
	5	0.5	1.0	0.0	—	—
	6	0.488	0.834	0.142	—	—
“vira”/“myna”	1	0.519	—	—	0.314	0.724
	2	0.491	—	—	0.230	0.752
	3	0.511	—	—	0.320	0.702
	4	0.469	—	—	0.747	0.191
	5	0.516	—	—	0.157	0.876
	6	0.522	—	—	0.140	0.904
“up”/“myna”	1	0.912	0.885	—	—	0.938
	2	0.971	0.942	—	—	1.0
	3	0.764	0.965	—	—	0.567
	4	0.736	0.931	—	—	0.544
	5	0.957	0.982	—	—	0.932
	6	0.815	0.834	—	—	0.797
“up”/“vira”	1	0.974	0.948	—	—	1.0
	2	0.798	0.76	—	—	0.837
	3	0.929	0.977	—	—	0.882
	4	0.915	0.937	—	—	0.893
	5	0.957	0.931	—	—	0.983
	6	0.824	0.851	—	—	0.797
“down”/“vira”	1	0.541	—	0.16	0.915	—
	2	0.940	—	0.88	1.0	—
	3	0.898	—	0.885	0.910	—
	4	0.886	—	0.908	0.865	—
	5	0.719	—	0.977	0.466	—
	6	0.940	—	0.88	1.0	—
“down”/“myna”	1	0.971	—	0.942	—	1.0
	2	0.971	—	0.942	—	1.0
	3	0.895	—	0.96	—	0.831
	4	0.838	—	0.88	—	0.797
	5	0.900	—	0.965	—	0.837
	6	0.818	—	1.0	—	0.640

Table 6. Mean and median accuracy scores for Vision Transformer and ResNext-18

words pair	mean acc. Vision Transformer	mean acc. ResNext-18	median acc. Vision Transformer	median acc. ResNext-18
“up”/“down”	0.493	0.505	0.491	0.5
“vira”/“myna”	0.505	0.496	0.514	0.5
“up”/“myna”	0.859	0.694	0.864	0.706
“up”/“vira”	0.899	0.702	0.922	0.745
“down”/“vira”	0.821	0.732	0.892	0.688
“down”/“myna”	0.899	0.734	0.898	0.781

connection between statistical properties of the classes in the dataset and the behavior of the classifiers. In the process of forming the criterion, we will use only the p-values computed for training and validation parts of the dataset as the test part of the dataset normally remains hidden from the researcher. Consider a pair of words i, j. Let's denote p_{tr} as the p-value obtained from the Kolmogorov – Smirnov test for words i and j on the training sets, p_{val} as the p-value on the validation sets. Then the following estimates hold for the median accuracy a of the binary classification of words i, j:

1) if $p_{tr} \geqslant 5 \times e^{-1}$, then accuracy $\geqslant 0.5$;
2) if $5 \times e^{-1} > p_{tr} \geqslant 4.9 \times e^{-168}$, then accuracy $\geqslant 0.69$;
3) if $p_{tr} < 5 \times e^{-168}$ & $p_{val} \geqslant 1.5 \times e^{-52}$, then accuracy $\geqslant 0.71$;
4) if $p_{tr} < 4.9 \times e^{-168}$ & $p_{val} < 1.5 \times e^{-52}$, then accuracy $\geqslant 0.78$.

It is worth noting that the stricter the conditions on the range of p-values, the greater the lower bound of the accuracy score obtained by ResNext-18 network. This fact implies that there exists a correlation between p-values obtained from the Kolmogorov – Smirnov method application and the accuracy of the network classifier trained on the examined data.

Predictive criterion for Vision Transformer

We further describe an analogous predictive criterion for the case of Vision Transformer classifier. As in the previous paragraph, we consider a pair of words i, j. Let's denote p_{tr} as the p-value obtained from the Kolmogorov – Smirnov test for words i and j on the training sets, p_{val} as the p-value on the validation sets. Then the following estimates hold for the median accuracy a of the binary classification of words i, j:

1) if $p_{tr} \geqslant 4.99 \times e^{-1}$, then accuracy $\geqslant 0.49$;
2) if $4.9 \times e^{-1} > p_{tr} > 4.9 \times e^{-168}$ or $p_{val} > 9.9 \times e^{-58}$, then accuracy $\geqslant 0.86$;
3) if $p_{tr} \leqslant 4.9 \times e^{-168}$ & $p_{val} \leqslant 9.9 \times e^{-58}$, then accuracy $\geqslant 0.89$.

The suggested criterion describes the classification behavior of Vision Transformer and holds the following condition: stricter bounds on p-values result in improved lower estimation bound on accuracy score.

General criterion for Vision Transformer and ResNext-18

The general criterion which is presented further is true for both classifiers considered and allows one to predict the results of both networks with an estimate from below for classification accuracy as follows:

1) if $p_{tr} \geqslant 4.99 \times e^{-1}$, then accuracy $\geqslant 0.49$;
2) if $4.9 \times e^{-1} > p_{tr} \geqslant 4.9 \times e^{-168}$, then accuracy $\geqslant 0.69$;
3) if $p_{tr} < 4.9 \times e^{-168}$ & $p_{val} \geqslant 1.5 \times e^{-52}$, then accuracy $\geqslant 0.71$;
4) if $p_{tr} < 4.9 \times e^{-168}$ & $p_{val} < 1.5 \times e^{-52}$, then accuracy $\geqslant 0.78$.

The general criterion is almost similar to the criterion suggested for ResNext-18 with the exception of the area described with the equation $p_{tr} \geqslant 4.99 \times e^{-1}$. This phenomenon is due to the fact that the ResNext-18 accuracy score is lower than the score of Vision Transformer in each of the remaining areas, so that the lower bound on accuracy score is set by ResNext-18. In the case of the p-value area $p_{tr} \geqslant 4.99 \times e^{-1}$, Vision Transformer performs slightly worse than ResNext-18, setting the lower accuracy bound 0.49 in this zone.

Discussion

It can be noticed that the Kolmogorov – Smirnov test applied to the word pairs "up"/"down" and "vira"/"myna" results in significantly higher p-values than any other word pairs. Indeed, the minimum p-value computed among all three components for the word pair "up"/"down" equals 0.161 (achieved in the test dataset) and for the word pair "vira"/"myna" — 0.421 (achieved in the validation dataset). At the same time, all other p-values presented in Tables 1–3 computed for the remaining word pairs, do not exceed $1.11 \times e^{-16}$. This observation allows us to conclude that regarding EEG signal distribution, the word "up" is statistically proximate to the word "down", and the word "vira" is statistically proximate to the word "myna".

One possible reason for such a phenomenon is that the above-mentioned pairs consist of semantically close words which describe opposite directions of movement in the vertical plane. The words "up" and "down" are commonly used in everyday life, while the words "vira" and "myna" are really narrow-specialized despite having the same motional direction meaning.

At the same time, the Kolmogorov – Smirnov test shows that mixing a common word and an uncommon one to form a word pair results in significantly lower levels of p-values. In other words, classes representing wide-spread words (such as "up" and "down") are statistically remote from classes representing unfrequent words (e. g., "vira", "myna") and, therefore, may lead to higher accuracy scores of neural network classifiers.

The binary classification task results show that both word pairs "up"/"down" and "vira"/"myna" cannot be separated well by ResNext-18. According to the set of p-values obtained from the Kolmogorov – Smirnov test execution, we claim that empirical statistical distributions of the corresponding data sets coincide. This is the main reason of the poor classification behavior of the neural network model. Indeed, it can be seen from Table 4 that in most experiments, either the classifier is confused when trying to classify tensors belonging to classes "up" and "down". As a result, either all tensors from the class "up" are referred to class "down" (leading to 0.0 accuracy score for class "up" and 1.0 score for class "down"), or vice versa. The same situation arises when distinguishing between the words "vira" and "myna". ResNext-18 either classifies all tensors in the dataset as a single class or makes random guesses, which results in 0.5 accuracy score for both classes. However, the situation is the opposite of the remaining word pairs, which correspond to statistically separable classes (i. e., with low p-values). The classifier tends to distinguish well between the classes in the majority of cases, leading to significantly higher levels of test accuracy scores (0.7–0.8).

We see that different network architectures "behave" similarly in relation to this task. We can see that the Vision Transformer gave us the same regularities in the set of test accuracy scores as in the case of ResNext-18. However, the overall level of accuracy is higher than for ResNext-18.

In the "Statistical predictive criterion" section, we introduced the statistical criterion, which establishes the connection between the p-values obtained from the Kolmogorov – Smirnov test and the accuracy scores of trained classifiers. We formed the predictive criterion which fits both ResNext-18 and Vision Transformer architectures simultaneously to prove the hypothesis stated in the introduction.

Overall, we see an inverse correlation between the p-value and the classification accuracy of the neural networks. Based on the results obtained, we can describe all cases of classification behavior depending only on the results of the Kolmogorov – Smirnov test and form the following predictive criteria for Vision Transformer and ResNext-18.

The correlation between p-values obtained from the Kolmogorov – Smirnov method application and the accuracy of the network classifier trained on the examined data proved the hypothesis for the ResNext-18 and Vision Transformer case. Furthermore, the described criterion may be used for further ResNext-18 and Vision Transformer classification behavior prediction.

Conclusion

In this work, we have shown the fundamental possibility of working with low-dimensional data and reducing resource consumption to select the optimal BCI dictionary based on EEG data and silent speech recognition. We formulated and confirmed the hypothesis about the existence of a predictive criterion based on distribution statistics of data with reduced dimension and predicting the accuracy of classification by a neural network on the same data of a higher dimension. The predictive criterion sets a connection between the accuracy score of a given classifier and the statistical properties (i. e., statistical separability, which is quantitatively described by p-values) of the classes in the dataset. We also showed that higher levels of *p*-values between the classes lead to the degradation of classification behavior. In particular, by applying the data processing approach proposed in this paper we proved that a dataset consisting of 2 statistically remote words yields significantly higher accuracy metrics in the problem of binary classification than a dataset with two statistically close words. The application of the Kolmogorov – Smirnov test showed us that electrical patterns produced by the silent pronunciation of words "up" and "down" share the same statistical distribution. At the same time, electrical activity signals for the word "vira", a synonym for "up", form a completely different distribution from the word "down". We showed that both of the neural network classifiers used in the research reached only 50 % accuracy score on the out-of-sample test dataset formed from the words "up" and "down", corresponding to the baseline accuracy level, while reaching up to 89 % score on the test set consisting of the words "vira" and "down". The highest value of accuracy metrics is obtained when classifying the words "up" and "vira" and equals 92 %. The idea underlying the results presented in this paper can be further expanded for the problem of multi-class classification, getting us closer to creating a dictionary of arbitrary size, including only semantically and phonetically remote words.

The criterion presented in this article has its limitations. Until there is a theoretical proof of a generalized criterion, we cannot apply this criterion to unknown network architectures and other types of classifiers.

References

Aarsland D., Bernadotte A. Epidemiology of dementia associated with Parkinson's disease // *Emre M.* (ed.) Cognitive impairment and dementia in Parkinson's disease. — 2 ed. — Oxford, 2015. — https://doi.org/10.1093/med/9780199681648.003.0002 (accessed: 9.04.2023).

Balaji A., Haldar A., Patil K., Ruthvik T. S., Valliappan C. A., Jartarkar M., Baths V. EEG-based classification of bilingual unspoken speech using ANN / Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). — Jeju, Korea, 2017.

Bernadotte A. Algorithm maximizing precision of *k*-classification on representatives of *k* equivalence classes for applied problem of selection of command dictionary / Academician O. B. Lupanov XIV International Scientific Seminar "Discrete Mathematics and Its Applications" (20–25 June 2022, Moscow). Keldysh Institute of Applied Mathematics. — 2022a. — P. 177–180. — doi.org/10.20948/dms-2022-53

Bernadotte A. The Algorithm that maximizes the accuracy of *k*-classification on the set of representatives of the *k* equivalence classes // Mathematics. — 2022b. — Vol. 10, No. 15. — P. 2810. — DOI: https://doi.org/10.3390/math10152810

Bernadotte A., Mikhelson V., Spivak I. Markers of cellular senescence. Telomere shortening as a marker of cellular senescence // Aging (Albany NY). — 2016. — Vol. 8. — P. 3–11. — DOI:10.18632/aging.100871

Bernadotte A., Mikhelson V. M., Spivak I. M. et al. Influence of donor age on cellular ability to carry out DNA repair via homologous recombination // Adv. Gerontol. — 2014. — Vol. 4. — P. 171–175. — https://doi.org/10.1134/S2079057014030023

Brigham K., Kumar B. Imagined speech classification with EEG signals for silent communication: A preliminary investigation into synthetic telepathy / Proceedings of the 2010 4th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2010. — Chengdu, China, 2010.

Bromley J., Bentz J. W., Bottou L., Guyon I., LeCun Y., Moore C., Säckinger E., Shah R. Signature verification using a "siamese" time delay neural network. Advances in neural information processing systems // Int. J. Pattern Recognit. Artif. Intell. — 1994. — Vol. 7. — P. 669–688.

Cooney C., Folli R., Coyle D. Mel frequency cepstral coefficients enhance imagined speech decoding accuracy from EEG / Proceedings of the 2018 29th Irish Signals and Systems Conference (ISSC). — Belfast, UK, 2018.

DaSalla C. S., Kambara H., Sato M., Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns // Neural Netw. — 2009. — Vol. 22. — P. 1334–1339.

Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An image is worth 16×16 words: Transformers for image recognition at scale // arXiv. — 2020. — arXiv:2010.11929.

Huang G.-B., Zhu Q.-Y., Siew C.-K. Extreme learning machine: A new learning scheme of feedforward neural networks / Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). — Budapest, Hungary, 2004.

Karpenko M. N., Vasilishina A. A., Gromova E. A., Muruzheva Z. M., Bernadotte A. Interleukin-1b, interleukin-1 receptor antagonist, interleukin-6, interleukin-10, and tumor necrosis factor-alpha levels in CSF and serum in relation to the clinical diversity of Parkinson's disease // Cellular Immunology. — 2018. — Vol. 327. — P. 77–82. — DOI:10.1016/j.cellimm.2018.02.011

Kuchaiev O., Boris G. Training deep AutoEncoders for collaborative filtering // arXiv. — 2017. — arXiv:1708.01715

Lebedev M. Decoding movements from cortical ensemble activity using a long short-term memory recurrent network // Neural Computation. — 2019. — Vol. 31, No. 6. — P. 1085–1113.

Mazurin A., Bernadotte A. Clustering quality criterion based on the features extraction of a tagged sample with an application in the field of brain-computer interface development // Intelligent systems. Theory and Applications. — 2021. — Vol. 25, No. 4. — P. 322–327. — http://intsysjournal.org/pdfs/25-4/MazurinBernadott.pdf

Min B., Kim J., Park H. J., Lee B. Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram // Biomed. Res. Int. — 2016. — 2618265.

Nguyen C. H., Karavas G. K., Artemiadis P. Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features // J. Neural Eng. — 2018. — Vol. 15. — 016002.

Ossadtchi A., Lebedev M. Consensus on the reporting and experimental design of clinical and cognitive-behavioral neurofeedback studies // Brain. — 2020. — Vol. 143, No. 6. — P. 1674–1685.

Panachakel J. T., Ramakrishnan A. G., Ananthapadmanabha T. V. Decoding imagined speech using wavelet features and deep neural networks / Proceedings of the 2019 IEEE 16th India Council International Conference (INDICON). — Rajkot, India, 2019.

Pandey A., Wang D. TCNN: temporal convolutional neural network for real-time speech enhancement in the time domain / Proceedings of the ICASSP 2019 — 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — Brighton, UK, 2019.

Pramit S., Muhammad A.-M., Sidney F. SPEAK YOUR MIND! Towards imagined speech recognition with hierarchical deep learning // arXiv. — 2019. — arXiv:1904.04358

Vasilishina A., Kropotov A., Spivak I., Bernadotte A. Relative human telomere length quantification by real-time PCR / *Demaria M.* (ed.) Cellular senescence. Methods in molecular biology. — Vol. 1896. — New York, NY: Humana Press, 2019. — https://doi.org/10.1007/978-1-4939-8931-7_5

Vorontsova D., Menshikov I., Zubov A., Orlov K., Rikunov P., Zvereva E., Flitman L., Lanikin A., Sokolova A., Markov S., Bernadotte A. Silent EEG-speech recognition using convolutional and recurrent neural network with 85 % accuracy of 9 words classification // Sensors. — 2021. — Vol. 21, No. 20. — P. 6744. — DOI: https://doi.org/10.3390/s21206744

Xie S., Girshick R., Dollár P., Tu Zh., He K. Aggregated residual transformations for deep neural networks // arXiv. — 2016. — arXiv:1611.05431.

Zubov A., Isaeva M., Bernadotte A. Neural network classifier of EEG-data from people who have undergone COVID-19 and have not // Intelligent systems. Theory and Applications. — 2021. — Vol. 25, No. 4. — P. 318–322. — http://intsysjournal.org/pdfs/25-4/ZubovIsaevaBernadott.pdf