

## МЕТОДЫ МАШИННОГО РАСПОЗНАВАНИЯ КОМАНД В ЭЭГ-СИГНАЛЕ

Буданцев А. В., Скляр А. Я.

*МИРЭА – Российский технологический университет, 119454, Россия, г. Москва, проспект Вернадского, 78, e-mail: budancev@mirea.ru, askliar@mail.ru*

---

В статье рассмотрены методы анализа электроэнцефалографического сигнала (ЭЭГ-сигнала) с целью выделения в нем характерных признаков для их последующей классификации. Исследуется возможность классификации ЭЭГ-данных о движении двумя испытуемыми различными группами мышц лица. Для получения электроэнцефалограмм и их передачи для обработки использовалась аппаратура Emotiv EPOC+ с беспроводным интерфейсом подключения к компьютеру и возможностью 14-канальной записи. Анализ данных проводился по результатам 150 сессий по 16 секунд в каждой, в рамках которых рассматривалось 5 различных реакций. В работе рассматриваются и сравниваются различные способы извлечения признаков из ЭЭГ-сигнала, а именно: коэффициенты спектрального разложения Фурье, значения функции Альтера-Джонсона, SAX-BOP признаки, линейные корреляции и классические статистики. Тестирование проводится в рамках двух нейросетевых моделей: сверточной и полносвязанной сетей. Кроме того, рассмотрены также методы на основе прямого анализа статистической информации. По результатам исследования получены оценки точности выявления тестируемых реакций при различных способах извлечения признаков. Лучшие результаты выявлены для методов на основе спектрального разложения Фурье на полносвязанные сети – порядка 95% и 90% на обучающей и тестовой выборках и с использованием значений функции Альтера-Джонсона на сверточной сети – порядка 99% и 86%.

---

Ключевые слова: электроэнцефалограмма, машинное обучение, нейронная сеть, анализ данных, распознавание, временной ряд, интерфейс мозг-компьютер.

## METHODS OF COMPUTER RECOGNITION OF COMMANDS IN EEG-SIGNAL

Budancev A. V., Skliar A. J.

*MIREA - Russian Technological University, 119454, Moscow, 78 Vernadskogo Avenue, Russia, e-mail: budancev@mirea.ru, askliar@mail.ru*

---

The article discusses some methods of analyzing the electroencephalographic signal (EEG signal) in order to identify the characteristic features in it for their subsequent classification. The possibility of classifying movement EEG data by two subjects of different groups of facial muscles is investigated. To receive electroencephalograms and transmit them for processing, Emotiv EPOC+ equipment with a wireless interface for connecting to a computer and the possibility of 14-channel recording was used. The data analysis was based on the results of 150 sessions of 16 seconds each, in which 5 different reactions were considered. The paper considers and compares various methods of extracting features from an EEG signal, namely: the coefficients of the Fourier spectral decomposition, the values of the Alter-Johnson function, SAX-BOP features, linear correlations, and classical statistics. Testing is carried out in the framework of two neural network models: convolutional and fully connected networks. In addition, methods based on direct analysis of statistical information are also considered. According to the results of the study, estimates of the accuracy of detecting the tested reactions with various methods of extracting features were obtained. The best results were found for methods based on the spectral Fourier decomposition on a fully connected network-about 95% and 90% on the training and test samples, and using the values of the Alter-Johnson function on the convolutional network-about 99% and 86%.

---

Key words: electroencephalogram, machine learning, neural network, data analysis, recognition, time series, brain-computer interface.

## **Введение**

Проблема совершенствования человеко-машинных интерфейсов на сегодняшний день является актуальной, ведь сегодня компьютеры стали неотъемлемой частью нашей жизни. Вычислительные системы постоянно развиваются, растут объемы памяти и быстродействие, что создает новые возможности взаимодействия с ними.

В данной работе рассматривается возможность использования вычислительных систем для обработки данных об активности головного мозга вблизи кожи головы. Такие системы и алгоритмы создают основу для ускорения и облегчения повседневной работы с компьютером и упрощают доступ к нему различным людям.

Для реализации поставленных задач применяется нейрокомпьютерный интерфейс (далее НКИ), под которым понимается система (программно-аппаратный комплекс), предназначенный для прямого обмена информацией между мозгом и электронным устройством, снимающую информацию непосредственно с мозга.

На данный момент существуют однонаправленные и двунаправленные НКИ. Первые используются только для передачи информации от мозга или только для принятия информации в мозг. Двунаправленные интерфейсы используются для приема/передачи информации в обе стороны. В данной статье рассматриваются однонаправленные интерфейсы для передачи информации от мозга в электронное устройство.

НКИ также различают по способам подключения к мозгу. По этому признаку они делятся на:

- Инвазивные – датчики мозговой активности вживляются непосредственно в головной мозг.
- Полуинвазивные – датчики находятся под черепной коробкой, но без непосредственного контакта с клетками мозга.
- Неинвазивные – датчики находятся на поверхности кожи головы или даже несколько удалены от нее.

Инвазивные и полуинвазивные виды НКИ обладают большей точностью фиксирования мозговой активности, но требуют хирургических вмешательств для подключения, что затрудняет исследования. Поэтому в данной работе рассматриваются неинвазивные НКИ.

Наиболее доступными однонаправленными неинвазивными НКИ на сегодня являются устройства, снимающие колебания электромагнитного поля с поверхности кожи головы при помощи электродов -- энцефалографы.

На рынке представлено множество подобных устройств. В рамках данной работы для анализа существующих алгоритмов было выбрано устройство Emotiv EPOC+ от австралийской компании Emotiv Systems. Оно было представлено в 2009 году, имеет беспроводной интерфейс подключения к компьютеру, поставляется с SDK, имеет возможность 14-канальной записи.

Принцип работы устройства основан на регистрации биоэлектрических сигналов, источниками которых могут быть отдельные клетки и связи между ними в головном мозге. Каждый раз, когда человек совершает ту или иную умственную работу, клетки мозга (нейроны) начинают обмениваться сигналами друг между другом. Сигналы, формируются за счет перемещения электрических зарядов между связанными нейронами. При перемещении зарядов часть энергии излучается в виде электромагнитных волн. Их и регистрирует устройство с частотой 128 Гц, после чего усиливает, оцифровывает и передает в компьютер. Стоит отметить также, что прямо на устройстве происходит частотная фильтрация с полосой пропускания 0.2 – 45 Гц для минимизации влияния шумов на качество сигнала.

### **Данные для анализа**

Для целей данного исследования были записаны данные с двух испытуемых. В целях уменьшения помех к испытуемым принимались требования минимальности волосяного покрова (волосы длиной не более сантиметра). Запись производилась следующим образом. На голову испытуемых надевалось устройство для записи ЭЭГ-данных – Emotiv EPOC+. Электроды устройства смачивались электролитическим раствором. Далее устройство включалось, соединялось с компьютером по беспроводной связи, и начинался сеанс записи данных. Во время сеанса испытуемые находились в расслабленном состоянии в положении сидя перед экраном монитора. На последнем отображалось действие, которое испытуемый в данный момент времени должен выполнить. Всего таких действий было пять: напряжение левой щеки, напряжение правой щеки, поднятие бровей вверх, опускание бровей вниз и отдых. Сеанс записи делился на сессии по 16 секунд. Во время одной сессии записывался ровно один вид действия. Всего было записано по 20 сессий на каждое из пяти действий для одного испытуемого и 10 сессий на каждое из пяти действий для другого. Итого в сумме 150 сессий по 16 секунд. Процесс записи данных, деление его на сессии, а также разметка данных производилась с помощью, разработанной специально для исследования, программы, интерфейс которой представлен на рисунке 1.

Фрагменты записи результатов измерений по отдельным каналам для разных классов данных представлены на рисунке 2. Именованию каналов соответствует международная система 10 – 20. По оси ординат на всех графиках значение разности потенциалов (в мВ) между референтным электродом и соответствующим каналу. По оси абсцисс номер отсчета.

### **Подготовка данных к обработке.**

Для выявления признаков и классификации удобно делить временные ряды на сегменты одинаковой продолжительности для изоляции отличительных черт друг от друга. Такие сегменты так же являются временными рядами, и мы далее будем именовать их окнами.

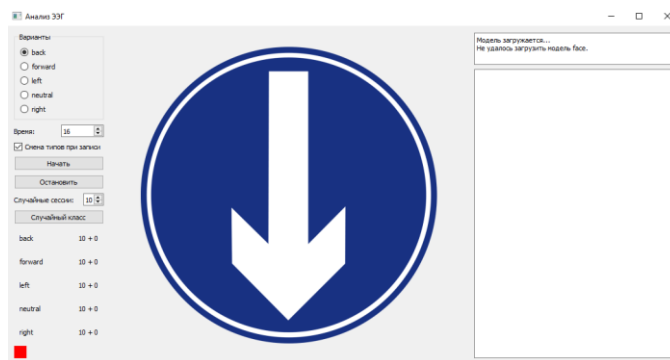


Рисунок 1. Программный интерфейс записи данных

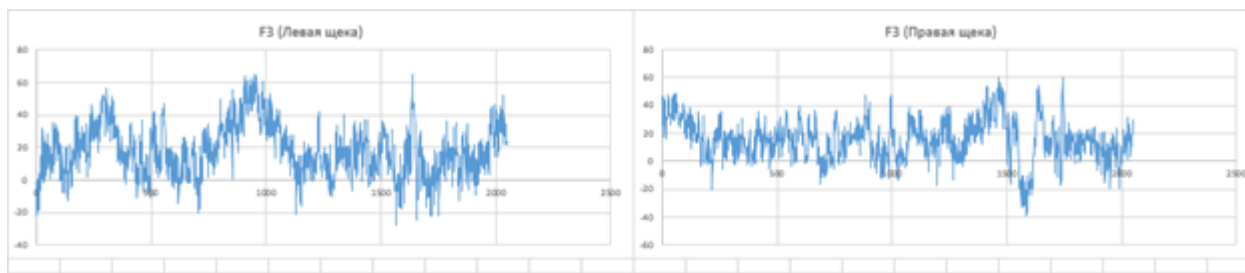


Рисунок 2. Записанные данные

Для целей данной работы размер окна составлял 128 отсчетов, что соответствует одной секунде записи. При этом окна брались со сдвигом в 16 отсчетов. Это значит, что одни и те же отсчеты попадали в несколько окон одновременно. Такая практика позволяет увеличить количество данных для обучения алгоритмов и построения более качественных моделей [1]. Таким образом, после деления на окна исходные данные можно представить в виде:

$$\{(\mathbf{X}^{(1)}, y^{(1)}), (\mathbf{X}^{(2)}, y^{(2)}), \dots, (\mathbf{X}^{(m)}, y^{(m)})\} \quad (1)$$

Где  $m$  – количество окон,  $\mathbf{X}$  – матрица отсчетов в окнах для каждого канала,  $y$  – целое число (0 - 4), означающее метку класса действия.

#### Методы извлечения признаков и используемые модели

Для классификации полученных окон по выполняемым действиям необходимо произвести расчет признаков. Под признаками понимается набор числовых значений, характеризующих классифицируемый объект (в данной случае матрицу с отсчетами окон). Чем лучше разделяются объекты в пространстве признаков, тем точнее может быть проведена классификация [2].

В данной работе рассматриваются следующие методы извлечения признаков:

1. Частотный спектр Фурье
2. Сдвиговая функция Альтера-Джонсона
3. SAX-WOP преобразование
4. Линейные корреляции окон
5. Статистики по окнам

После извлечения признаков происходит обучение алгоритма на этих признаках с целью получения модели зависимости условной вероятности появления метки определенного класса от набора признаков. В рамках данной работы рассматривались нейросетевые подходы к построению требуемых моделей, а именно модели на основе полносвязной и сверточной нейронных сетей. Рассмотрим подробнее суть каждого из названных методов.

#### Частотный спектр Фурье

Для получения частотного спектра используют преобразование Фурье. Так как исследуемый сигнал дискретен, удобно использовать Быстрое Преобразование Фурье (БПФ) [3]. Для лучшей работы алгоритм требует окна с количеством отсчетов, являющимся степенью двойки. Это частично обуславливает выбранный размер окна в 128 отсчетов.

В результате выполнения БПФ и взятия модуля получается набор чисел, пропорциональных мощности колебаний на определенных частотах. Так как частота дискретизации сигнала равна 128 Гц, а количество отсчетов в окне равно 128, то шаг частоты  $\frac{128}{128} = 1$  Гц. Так как исходный сигнал вещественный, то половина полученных чисел не несет информации и может быть отброшена согласно [3]. Таким образом, каждое окно преобразуется в набор из 64-х чисел. Далее для улучшения численной устойчивости от полученных чисел брался натуральный логарифм.

Эффективность использования коэффициентов частотного спектра в качестве признаков может быть обусловлена тем, что разная мозговая активность сопровождается разными ритмами [4]. Например, в моменты расслабления при закрытии глаз у большинства здоровых людей в затылочной доле на ЭЭГ фиксируется низкочастотный альфа-ритм, а при открытии глаз или концентрации высокочастотный бета-ритм. Кроме того, при совершении движений или представлении совершения движений пропадает мю-ритм в височных долях. Этот факт часто используется в НКИ [5]. По оси ординат откладывается мощность сигнала, по оси абсцисс – частота.

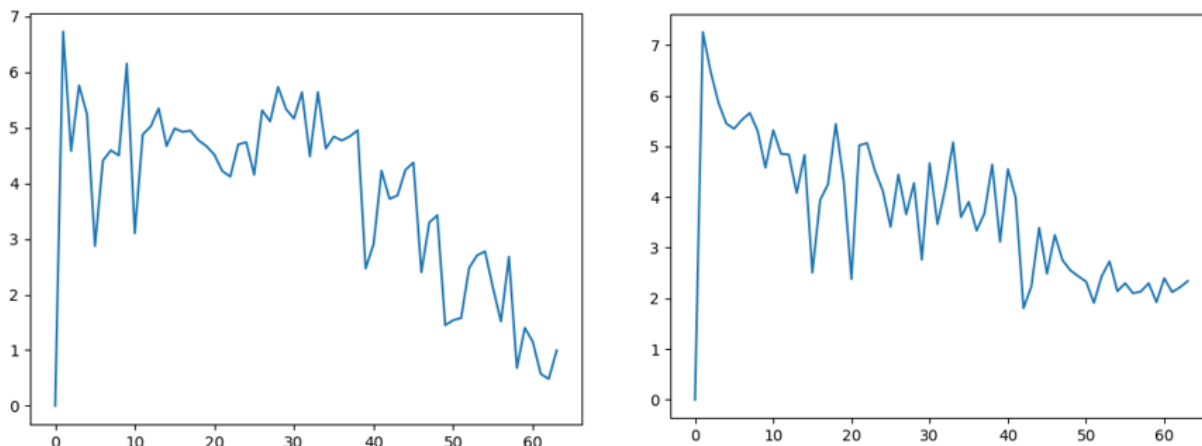


Рисунок 3. Варианты преобразования Фурье.

#### Сдвиговая функция Альтера-Джонсона

Функция Альтера-Джонсона позволяет найти периоды или почти-периоды в данных [6][7]. Рассчитывается она следующим образом:

$$a(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} |F_{t+\tau} - F_t|$$

Где  $F$  – исходный временной ряд,  $N$  – общее количество отсчетов во временном ряде.

Точки локальных минимумов этой функции называют почти-периодом. Действительно, если функция в некоторой точке принимает значение 0, то эта точка соответствует точному периоду исходного временного ряда. Чем ближе значение функции Альтера-Джонсона в точке почти-периода к нулю, тем «лучше» почти-период. Для получения признаков вычисляются значения функции Альтера-Джонсона на множестве целых чисел от 0 до  $0,75N$  для каждого окна в данных.

Для удаления постоянной составляющей далее происходит численное приближение производной по двум точкам от функции Альтера-Джонсона:

$$a'_t \approx \frac{a_{t+h} - a_{t-h}}{2h}$$

Полученные производные и берутся за признаки. Эффективность расчета признаков таким способом может быть обусловлена также, как и для преобразования Фурье, ведь функция Альтера-Джонсона дает некоторое представление о периодах колебательной составляющей временного ряда. Кроме того, предполагается, что ритмы могут иметь не только форму синусоиды, но и любую другую форму. Преобразование Фурье не всегда может корректно выделить такого рода колебания, но с помощью функции Альтера-Джонсона, которая не зависит от характера колебаний это всегда можно корректно выполнить [6].

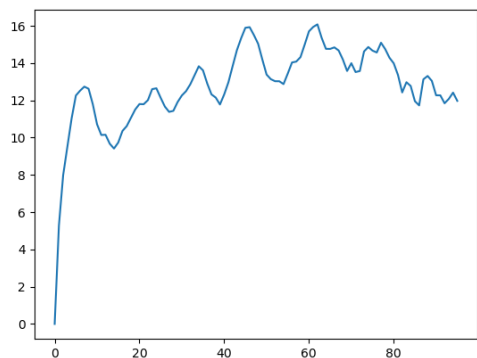


Рисунок 4. Функция Альтера-Джонсона на одном окне

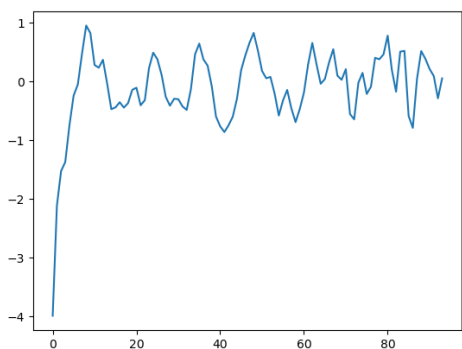


Рисунок 5. Приближение производной от функции Альтера-Джонсона

### SAX-BOP преобразование

Это преобразование также выполняется над каждым окном и состоит из двух этапов: SAX (Symbolic Aggregation approXimation) и BOP (Bag Of Patterns). Сначала область значений отсчетов в окне разбивается на равные по длине отрезки. Далее отсчеты, попавшие в разные отрезки, помечаются разными буквами или цифрами. Отсчеты в окне заменяются их метками. На этом заканчивается этап SAX. Далее происходит подсчет меток каждого типа и все окно заменяется полученными значениями. На этом заканчивается BOP.

Расчет признаков таким способом может быть полезен, так как, возможно, разные виды деятельности мозга могут сопровождаться определенными паттернами во временной области ЭЭГ-сигнала. При этом такого рода паттерны в разных окнах могут быть сдвинуты друг относительно друга. SAX-BOP преобразование вносит инвариантность относительно сдвига, что позволяет минимизировать влияние этого эффекта на классификацию.

### Линейные корреляции окон

Коэффициент корреляции Пирсона является классическим способом количественно оценить взаимосвязь между двумя случайными величинами. Он вычисляется следующим образом:

$$r_{xy} = \frac{M(x - M(x)) * M(y - M(y))}{\sigma_x \sigma_y}$$

На реальных данных математическое ожидание ( $M$ ) заменяется выборочным средним, а среднеквадратическое отклонение ( $\sigma$ ) – корнем квадратным из несмещенной оценки дисперсии.

В рамках данной работы рассчитывались коэффициенты корреляции между окнами (столбцами) в матрицах  $X$  из выражения (1). При этом каждая такая матрица заменялась набором корреляций, которые и брались за признаки.

Такой расчет признаков обусловлен тем, что активность разных зон мозга, особенно симметричных относительно линии, разделяющей полушария, может коррелировать во время отсутствия какой-либо деятельности. Во время напряжения мышц лица или другой активности, корреляции могут уменьшаться [8].

### Статистики по окнам.

Под статистиками в данном случае понимается два значения: среднее арифметическое значение отсчетов в окне и несмещенная оценка их дисперсии:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$
$$d = \frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Такой выбор способа расчета признаков обусловлен простотой визуализации окон в полученном пространстве и из предположения, что указанные статистики могут значительно меняться в зависимости от типа активности мозга.

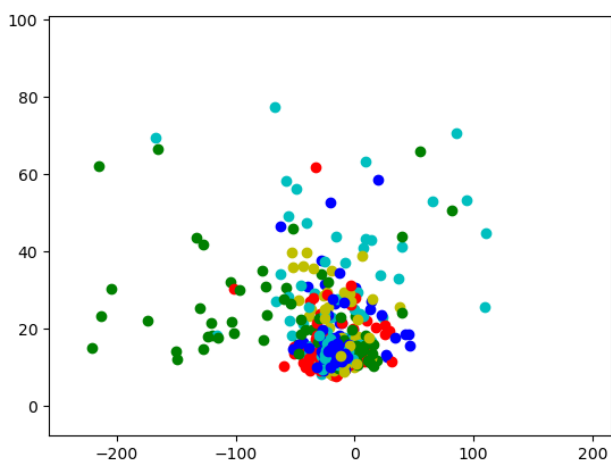


Рисунок 6. Среднее (ось абсцисс) и дисперсия (ось ординат) по окнам канала F3

Отметим, что при анализе может использоваться и расширенный набор статистических показателей, таких как мода, медиана, эксцессы и т.п.

### Описание моделей.

В качестве моделей в рамках данной работы рассматриваются искусственные нейронные сети. Они хорошо себя зарекомендовали в последнее время при решении многих задач. Основным полезным свойством нейронных сетей является их универсальность: они могут аппроксимировать почти любую зависимость при

достаточно удачной архитектуре [9]. В процессе исследования были рассмотрены две архитектуры: полносвязная нейронная сеть и сверточная.

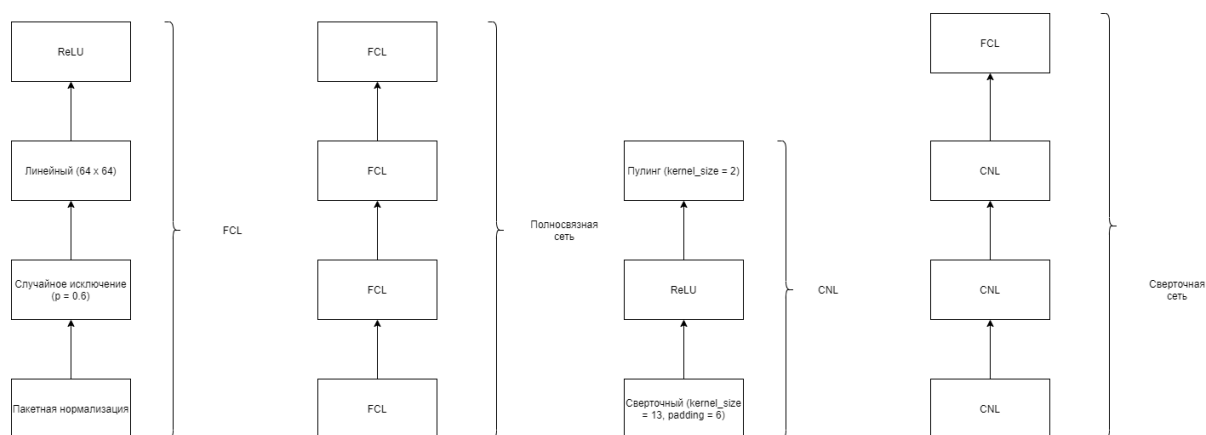


Рисунок 8. Архитектуры полносвязной и сверточной нейронных сетей

Слой пакетной нормализации (Batch Normalization) был описан в [10], слой случайного исключения (Dropout) – в [11], сверточные нейронные сети с пулингом были исследованы в [12]. Количество обучаемых параметров полносвязной сети – 407 464, сверточной сети – 177 782.

Данные в сверточную сеть поступали без изменений, в виде трехмерного тензора. Свертка происходила по отсчетам в окне. При передачи данных в полносвязную сеть из трехмерного тензора сначала конструировалась матрица путем соединения всех окон из матриц  $X$  в выражении (1) в строки результирующей матрицы.

Модели были обучены на трех наборах данных: данные с первого испытуемого, данные со второго испытуемого, данные с обоих испытуемых одновременно. Для обучения и тестирования каждый набор данных делился на две части: обучающую и тестовую. Тестовая часть составляла около 12,5% от всего набора данных. Обучение происходило по мини-пакетам размером 256 элементов. В качестве функции потерь для обеих сетей использовалась перекрестная энтропия. Оптимизатор – Adam. В каждом случае обучение длилось по 25 эпох. На выходе сетей были получены 5 чисел за счет того, что в линейном слое последнего FCL-слоя размерность матрицы 64x5. Номер максимального из этих чисел принимался за предсказанных класс, соответствующий действию испытуемого.

#### Результаты исследования.

Показатели точности (ассигасу) моделей на наборах данных представлены в таблице 1. Так как при вычислении линейных корреляций между окнами теряется размерность и тензор входных данных превращается в матрицу, сверточная сеть не была обучена на таких признаках. При вычислении статистик также не использовалась сверточная нейронная сеть, так как в окне становится слишком мало отсчетов (всего 2). Наборы, полученные с первого, второго и с обоих испытуемых обозначены, как «И1», «И2», «И12» соответственно.

Таблица 1. Результаты

Признаки		Модель		Сверточная сеть	Полносвязная сеть	
		Обучающий набор	Тестовый набор	И1	И2	И12
Спектр Фурье	Обучающий набор			И1	96,68%	98,4%
				И2	86,88%	91,99%
				И12	94,1%	93,65%
	Тестовый набор			И1	94,67%	95,68%
				И2	84,98%	89,06%
				И12	82,02%	91,98%
Функция Альтера-Джонсона	Обучающий набор			И1	99,77%	95,16%
				И2	98,26%	87,3%
				И12	99,47%	92,68%
	Тестовый набор			И1	89,83%	85,73%
				И2	81,26%	72,09%
				И12	89,42%	62,73%
SAX-ВОР преобразование	Обучающий набор			И1	46,6%	32,78%
				И2	53,08%	62,48%
				И12	38,28%	47,78%
			И1	29,07%	29,09%	

	Тестовый набор	И2	27,71%	34,66%
		И12	25,89%	31,66%
Линейные корреляции окон	Обучающий набор	И1	-	35,33%
		И2	-	51,42%
		И12	-	33,11%
	Тестовый набор	И1	-	23,21%
		И2	-	43,56%
		И12	-	26,3%
Статистики по окнам	Обучающий набор	И1	-	71,98%
		И2	-	61,23%
		И12	-	59,07%
	Тестовый набор	И1	-	69,71%
		И2	-	50,12%
		И12	-	46,14%

В результате данной работы были исследованы различные методы извлечения признаков из ЭЭГ-данных, а также протестированы две нейросетевые архитектуры на разных признаках. Усреднив точности моделей на разных наборах данных и отсортировав по убыванию средней точности результаты, можно понять, какие признаки с какими архитектурами сработали лучше. Эти результаты представлены в таблице 2 для сверточной нейронной сети и в таблице 3 для полносвязной.

Таблица 2. Признаки для сверточной нейронной сети

Признаки	Средняя точность на тестовой части
Спектр Фурье	87,22%
Функция Альтера-Джонсона	86,84%
SAX-BOP преобразование	27,56%
Линейные корреляции окон	-
Статистики по окнам	-

Таблица 3. Признаки для полносвязной нейронной сети

Признаки	Средняя точность на тестовой части
Спектр Фурье	92,24%
Функция Альтера-Джонсона	73,52%
Статистики по окнам	55,32%
SAX-BOP преобразование	31,8%
Линейные корреляции окон	31,02%

В рамках данного исследования полносвязная сеть в совокупности с признаками, основанными на спектральных характеристиках сигнала, показала лучший результат. Вопросом отдельного изучения остается подбор архитектуры сети для каждого из рассмотренных методом отдельно. Кроме того, можно рассмотреть гибридные методы вычисления признаков, использующие несколько методов сразу.

В заключение следует сказать, что предлагаемые методы анализа создают основу для выявления значительно более важных реакций, таких как: правда – ложь, гнев, радость, возбуждение, страх и т.д.

#### Список литературы

1. Коваленко П. Управление механическим манипулятором посредством нейрокомпьютерного интерфейса // Сборник тезисов XXIV Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2017». — Макс-Пресс, 2017.
2. Л. П. Попова, И. О. Датъев. Обзор существующих методов распознавания образов. — М.: Сборник научных трудов, 2007.
3. Нуссбаумер Г. Быстрое преобразование Фурье и алгоритмы вычисления сверток. — М.: Радио и связь, 1985.
4. Кирой В. Н., Ермаков П. Н. Общая характеристика ритмов ЭЭГ человека // Электроэнцефалограмма и функциональные состояния человека. — Ростов-на-Дону: Изд-во Рост. ун-та, 1998.
5. Pfurtscheller G., Christa N. EEG-Based Brain-Computer Interfaces // Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields / edited by D. L. Schomer, H. L. S. Fernando. — 6th. — Philadelphia, Pa.: Lippincott Williams & Wilkins, 2010. — P. 1227—1236. — 668 p.

6. Скляр А.Я. — Анализ временных рядов и выявление процессов с размытой периодичностью // Кибернетика и программирование. – 2018. – № 6. – С. 56 - 64.
7. Гадзаов А. Ф. — Алгоритмы выявления почти – периодов в результатах измерений // Журнал научных публикаций аспирантов и докторантов, ISSN 1991-3087
8. Кулаичев А.П. Метод анализа корреляционной синхронности ЭЭГ и его возможности. Журнал высшей нервной деятельности. 2011.
9. Cybenko, G. V. Approximation by Superpositions of a Sigmoidal function // Mathematics of Control Signals and Systems. — 1989. — Т. 2, № 4. — С. 303—314
10. Ioffe S., Szegedy C. — Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2016
11. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov —Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research 15, 2014.
12. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989.

---

## References

1. Kovalenko P. Control of a mechanical manipulator by neurocomputer interface // Collection of theses of the XXIV International Scientific Conference of students, postgraduates and young scientists "Lomonosov-2017". - Maks-Press 2017.
2. L. P. Popova, I. O. Datyev. Review of existing methods of pattern recognition. - M.: Collection of scientific papers, 2007.
3. Nussbaumer G. Fast Fourier transform and algorithms for calculating convolutions. - M.: Radio and Communications, 1985.
4. Kira V. N., Ermakov P. N. General characteristics of human EEG rhythms // Electroencephalogram and functional states of a person. - Rostov-on-Don: Publishing House of Rostov.
5. Pfurtscheller G., Krista N. Brain-computer interfaces based on EEG // Niedermeyer's electroencephalography: basic principles, clinical application and related areas / ed. by D. L. Schomer, H. L. S. Fernando.-6. - Philadelphia, PA: Lippincott Williams and Wilkins, 2010. - pp. 1227-1236 — - 668 p.
6. Sklyar A. Ya. - Analysis of time series and identification of processes with blurred periodicity // Cybernetics and Programming. – 2018. – No. 6. - pp. 56-64.
7. Gadzaov A. F.-Algorithms for detecting near-periods in measurement results // Journal of scientific publications of postgraduates and doctoral students-ISSN 1991-3087
8. Kulaichev A. P. Method of analysis of correlation synchronicity of EEG and its possibilities. Journal of Higher Nervous Activity. 2011.
9. Kibenko G. V. Approximation by superpositions of a sigmoidal function // Mathematics of control signals and Systems. - 1989. - Vol. 2, No. 4. - p. 303-314
10. Ioffe S., Szegedy S.-Batch normalization: Accelerating deep network learning by reducing the internal covariance shift, 2016
11. N. K. Srivastava, G. Hinton, A. Krizhevsky, I. Sutzkever, R. Salakhutdinov —Dropout: a Simple way to prevent overfitting in neural networks, Journal of Machine Learning Research 15, 2014.
12. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Back-propagation applied to the recognition of handwritten zip codes, Neural computation, 1(4):541-551, Winter 1989.