

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ
VIỆN TRÍ TUỆ NHÂN TẠO



**EEG Signal Classification for SSVEP-BCI: CCA
vs. UD M-CNN and UD C-CNN**

**BÁO CÁO MINI PROJECT
MÔN TÍNH TOÁN KHOA HỌC THẦN KINH VÀ ỨNG DỤNG -
AIT3012 37**

Sinh viên thực hiện:
Cao Xuân Nguyên - 22022670

TP. Hà Nội, 2024

Mục lục

Mục lục	2
1 Giới thiệu	1
2 Phương pháp	2
2.1 Canonical Correlation Analysis (CCA)	2
2.2 Convolutional Neural Network (CNN)	3
2.3 Magnitude Spectrum Features CNN (M-CNN)	3
2.4 Complex Spectrum Features CNN (C-CNN)	4
3 Thực nghiệm	5
3.1 Bộ dữ liệu và tiền xử lý	5
3.2 Triển khai mô hình và đánh giá	6
3.2.1 Huấn luyện mô hình	6
3.2.2 Kết quả và đánh giá	7
4 Kết luận	9
Tài liệu tham khảo	10

Chương 1

Giới thiệu

Hệ giao diện não-máy tính (Brain-Computer Interface, BCI) là một hệ thống cho phép giao tiếp trực tiếp giữa bộ não người và máy tính, thường thông qua việc ghi nhận và xử lý tín hiệu điện não (điện não đồ - EEG). BCIs cung cấp một con đường trực tiếp giữa não người và một thiết bị bên ngoài để giao tiếp và điều khiển mà không cần hệ thống thần kinh cơ thông thường. BCIs mở ra những khả năng mới cho việc phục hồi thần kinh đối với những người mắc các bệnh lý thần kinh như đột quỵ, xơ cứng teo cơ bên (ALS) hoặc liệt. Một hệ thống BCI có khả năng thu thập, phân tích và giải mã các thay đổi trong hoạt động não của người dùng, sau đó chuyển đổi chúng thành các lệnh để điều khiển hoặc tương tác với một thực thể bên ngoài. Thông thường một hệ thống BCI có năm giai đoạn liên tiếp như sau: thu nhận tín hiệu, tiền xử lý, trích xuất đặc trưng, phân loại và giao diện điều khiển hoặc ứng dụng.

Tác vụ Steady-State Visual Evoked Potentials (SSVEP) là một loại đáp ứng thị giác, được sinh ra khi người quan sát một kích thích nhấp nháy đều ở tần số cố định. Khi mắt tiếp nhận một kích thích nhấp nháy ở tần số cụ thể, tín hiệu EEG thu được sẽ có các thành phần dao động mạnh mẽ ở cùng tần số hoặc bội số của tần số kích thích. Đây là loại tín hiệu đặc biệt ổn định và dễ dàng đồng bộ với kích thích bên ngoài, cho phép nó được khai thác hiệu quả trong các ứng dụng BCI.

Tín hiệu SSVEP có những ưu điểm vượt trội như khả năng phân loại cao, không yêu cầu quá trình huấn luyện dài, và có thể được sử dụng trong các thiết lập BCI không xâm lấn. Trong nghiên cứu này, tôi tập trung vào việc phân tích so sánh khả năng phân loại tín hiệu EEG dựa trên ba phương pháp chính: Canonical Correlation Analysis (CCA), User-Dependent Magnitude spectrum CNN (UD M-CNN), và User-Dependent Complex Spectrum CNN (UD C-CNN) với CNN là kiến trúc mạng tích chập. Phương pháp User-Dependent bao gồm việc tập dữ liệu huấn luyện và mô hình được phát triển dựa trên các đặc điểm cụ thể của từng cá nhân, giúp khắc phục sự biến đổi giữa các đối tượng trong các hệ thống BCI.

Bài báo cáo tập trung nói về phương pháp và quá trình thực nghiệm và kết quả, mọi mã nguồn và kết quả chạy đều được lưu ở đường dẫn Github sau: [Tại đây](#)

Chương 2

Phương pháp

2.1 Canonical Correlation Analysis (CCA)

CCA là một kỹ thuật thống kê đa biến để xác định mối tương quan ẩn dưới tập hợp hai biến đa chiều. Việc xác định tương đương với việc tìm hai vectors cơ bản cho x và y sao cho độ tương quan (correlation) giữa phép chiếu của các biến đây lên các vector cơ bản đây là cực đại. Các nghiên cứu đã cho thấy CCA cho hiệu suất tốt trong việc phát hiện đáp ứng tín hiệu EEG SSVEP. Phương pháp này cũng được sử dụng rộng rãi như là Baseline cho các nghiên cứu liên quan đến phân loại đáp ứng SSVEP. Xét biến đổi tuyến tính $x = X^T w_x$ và $y = Y^T w_y$ với X tương ứng với tín hiệu EEG đa kênh và Y là một tập hợp tín hiệu tham chiếu (reference signals) có độ dài tương đương X . Mục tiêu CCA là tìm các vectors chiếu w_x và w_y để cực đại hoá tương quan của x và y như sau:

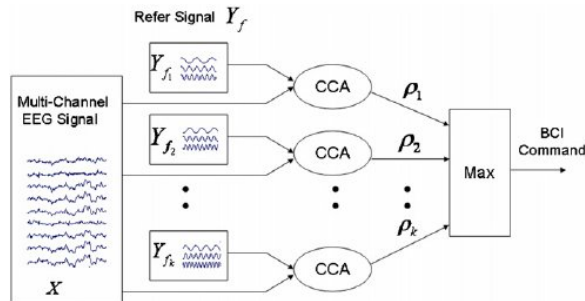
$$\rho(x, y) = \max_{w_x, w_y} \left[\frac{E[w_x^T X Y^T w_y]}{\sqrt{E[w_x^T X X^T w_x] \cdot E[w_y^T Y Y^T w_y]}} \right]$$

$$Y_n = \begin{bmatrix} \sin(2\pi f_n t) \\ \cos(2\pi f_n t) \\ \vdots \\ \sin(2\pi N_h f_n t) \\ \cos(2\pi N_h f_n t) \end{bmatrix}, \quad t = \left[\frac{1}{f_s}, \frac{2}{f_s}, \dots, \frac{N_s}{f_s} \right]$$

Với f_n là tần số kích thích, f_s là tần số lấy mẫu, N_s là số mẫu và N_h là số bội hài (harmonics). Giá trị tương quan tối đa ρ được tính toán cho từng phân đoạn tín hiệu EEG và lớp đầu ra C được xác định bằng:

$$C = \operatorname{argmax}(\rho_{f_i}) i = 1, 2, \dots, K.$$

Với K là số lớp. Trong tác vụ SSVEP, CCA như đã đề cập được sử dụng để tìm tần số kích thích tương ứng với tín hiệu EEG.



Hình 2.1: Minh họa việc phân tích tín hiệu EEG bằng CCA

2.2 Convolutional Neural Network (CNN)

Kiến trúc CNN là một kiến trúc phổ biến của mạng Neural sâu, chủ yếu dùng để phân loại đầu vào thành các nhãn xác định. Với việc áp dụng trên tác vụ SSVEP, CNN sẽ được sử dụng để phân loại tín hiệu đã được tiền xử lý theo hai hướng thành K nhãn ở lớp đầu ra bằng với số lớp SSVEP ở dữ liệu đầu vào. Ở báo cáo này sử dụng hai phương pháp để biến đổi miền dữ liệu cho lớp đầu vào của mạng CNN, còn kiến trúc chung được sử dụng như sau:

Layer	Dimensions	Description
Input	$N_{ch} \times N_{fc}$	Input features (N_{ch} channels, N_{fc} features per channel)
Conv_1	$2 \times N_{ch} \times 1 \times N_{fc}$	1D convolution across channels with kernel size $N_{ch} \times 1$
Conv_2	$2 \times N_{ch} \times 1 \times (N_{fc} - 10 + 1)$	1D convolution across frequency channels with kernel size 1×10
Batch Normalization	Same as input	Normalization of outputs from Conv_1 and Conv_2
ReLU	Same as input	Rectified Linear Unit activation function
Dropout	Same as input	Dropout regularization to prevent overfitting
Output	K	Softmax output layer with K units corresponding to SSVEP classes

Bảng 2.1: Kiến trúc được sử dụng của CNN

2.3 Magnitude Spectrum Features CNN (M-CNN)

CNN đã được trình bày ở trên như là một bộ phân loại tiềm năng cho dữ liệu EEG SSVEP. Các kiến trúc CNN có thể khác nhau ở dạng đầu vào của CNN, ví dụ như sử dụng các đặc trưng phổ biên độ (Magnitude Spectrum Features) làm đầu vào cho CNN (M-CNN [2]). Phương pháp này được đánh giá và so sánh với các phương pháp khác về khả năng tổng quát và giải mã các kích thích SSVEP thông qua khoảng cách giữa các xung (ISD) khác nhau. Trong phương pháp này, các đoạn tín hiệu EEG $x(n)$ trong miền thời gian đã được tiền xử lý để chuyển sang miền tần số $X(k)$ bằng biến đổi Fourier nhanh (Fast Fourier Transforms - FFT), hình thành một chuỗi phức dạng $Re(X_k) + jIm(X_k)$. Khi đó Phổ biên độ được tính theo công thức:

$$|X_k| = \sqrt{Re(X(k))^2 + Im(X(k))^2}$$

Độ phân giải tần số (Frequency resolution) của FFT được cố định là 0.2930 Hz và các thành phần tần số từ 3 Hz đến 35 Hz được chọn. Do đó, chiều dài của tín hiệu đã chuyển đổi bằng FFT là $N_{fc} = 110$. Vector đặc trưng tính toán dọc theo mỗi kênh được xếp chồng lại với nhau để tạo thành một ma trận có kích thước $N_{ch} \times N_{fc}$, trong đó N_{ch} là số lượng kênh và N_{fc} là số lượng thành phần tần số. Cuối cùng, ma trận này được cung cấp làm đầu vào cho CNN. M-CNN có kiến trúc CNN được nêu ở phần trước. Một ví dụ của đầu vào I_{M-CNN} có dạng:

$$\mathbf{I}_{M-CNN} = \begin{bmatrix} |FFT(x_{O1})| \\ |FFT(x_{Oz})| \\ |FFT(x_{O2})| \end{bmatrix}$$

Phương pháp này chỉ xem xét biên độ của tín hiệu tại các tần số khác nhau và bỏ qua thông tin pha. Việc bỏ qua thông tin này có thể gây ra hiệu suất kém trong việc giải mã tín hiệu

2.4 Complex Spectrum Features CNN (C-CNN)

Từ hạn chế của việc trích xuất đặc trưng phổ biến độ đã nêu, việc xem xét phương pháp có thể kết hợp thông tin về pha tín hiệu là cần thiết. Tín hiệu biểu diễn FFT được sử dụng để rút ra cả thông tin về biên độ và pha của tín hiệu.

Đầu tiên, tín hiệu trong miền thời gian chuyển đổi thành biểu diễn FFT sử dụng FFT chuẩn với độ phân giải 0.2930 Hz. Tiếp theo, các thành phần tần số của phần thực và phần ảo dọc theo mỗi kênh được trích xuất trong khoảng từ 3 Hz đến 35 Hz, tạo ra hai vectơ có độ dài 110 mỗi vector. Hai Vector này sau đó được nối lại thành một Vector đặc trưng duy nhất như sau: $I = Re(X)||Im(X)$ với nửa đầu chứa phần thực và nửa sau là phần ảo của tín hiệu phức FFT được biểu diễn trong vector dài 220. Vectơ đặc trưng kết quả được xếp chồng lại với nhau để tạo thành một ma trận có kích thước $N_{ch} \times N_{fc}$, trong đó $N_{fc} = 220$. Cách tiếp cận sử dụng FFT phức là đầu vào cho CNN là C-CNN[2]. Một ví dụ về đầu vào I_{C-CNN} có dạng:

$$\mathbf{I}_{C-CNN} = \begin{bmatrix} \text{Re}\{\text{FFT}(x_{O1})\}, \text{Im}\{\text{FFT}(x_{O1})\} \\ \text{Re}\{\text{FFT}(x_{Oz})\}, \text{Im}\{\text{FFT}(x_{Oz})\} \\ \text{Re}\{\text{FFT}(x_{O2})\}, \text{Im}\{\text{FFT}(x_{O2})\} \end{bmatrix}$$

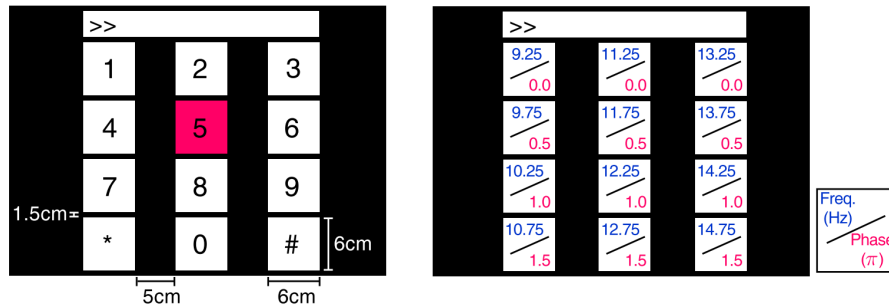
Chương 3

Thực nghiệm

3.1 Bộ dữ liệu và tiền xử lý

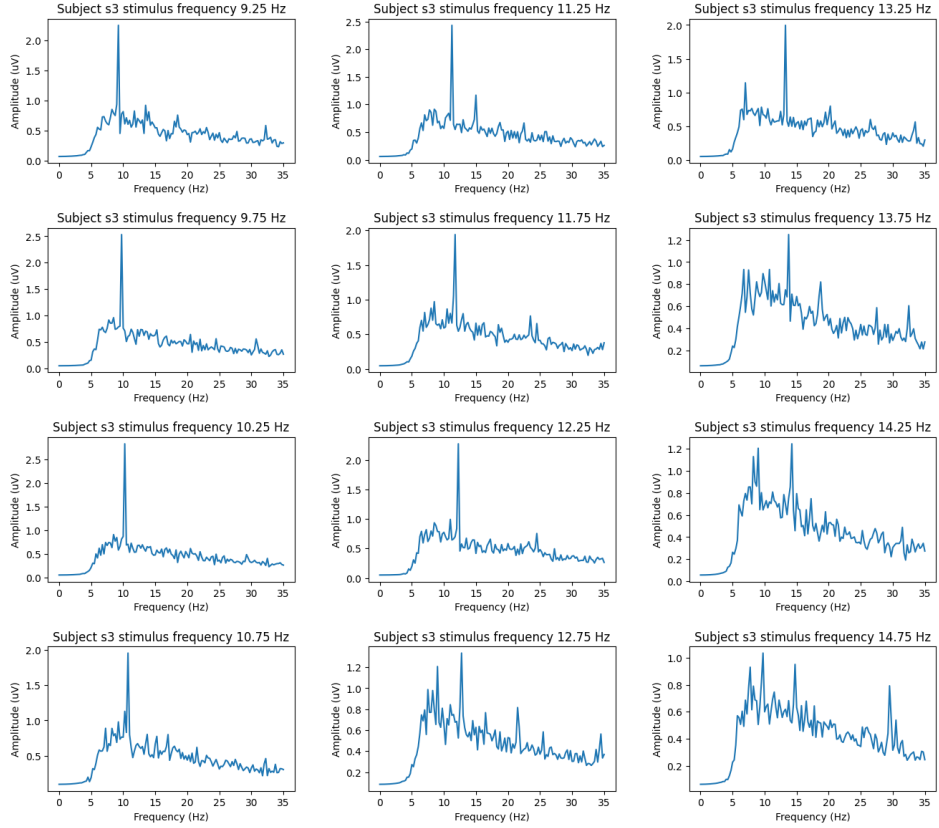
- Bộ dữ liệu: Bài thực nghiệm sử dụng bộ dữ liệu nguồn mở 12 lớp SSVEP được sử dụng trong nghiên cứu SSVEP BCI. Bộ dữ liệu được thu thập từ 10 người khỏe mạnh tham gia vào thí nghiệm với 12 kích thích thị giác nhấp nháy được trình bày trên màn hình. Tần số nhấp nháy dao động từ 9,25 Hz đến 14,75 Hz, với khoảng cách 0,5 Hz giữa các tần số. Các kích thích thị giác là các hình vuông 6×6 cm sắp xếp trong một ma trận 3×4 trên LCD đặt cách người tham gia 0,6 mét. Người tham gia tập trung vào một dấu cộng trung tâm trong 1 giây trước khi các kích thích bắt đầu nhấp nháy. Họ được yêu cầu tập trung vào mục tiêu đã được chỉ định trong suốt thời gian của mỗi thử nghiệm.

Các tín hiệu EEG được ghi lại sử dụng tám kênh. Mỗi người tham gia hoàn thành 15 khối, mỗi khối chứa 12 thử nghiệm (một thử nghiệm cho mỗi kích thích) được trình bày theo thứ tự ngẫu nhiên. Tổng cộng có 180 thử nghiệm cho mỗi người tham gia.



Hình 3.1: Mô tả thí nghiệm thu tín hiệu SSVEP EEG[1]

- Tiền xử lý: Tất cả tám kênh đã được sử dụng trong bộ dữ liệu này. Lọc theo thời gian được thực hiện bằng bộ lọc Butterworth băng thông 4 bậc với dải tần từ 6 Hz đến 80 Hz. Mỗi thử nghiệm dài 4 giây được chia thành các đoạn non-overlapping dài 1 giây.



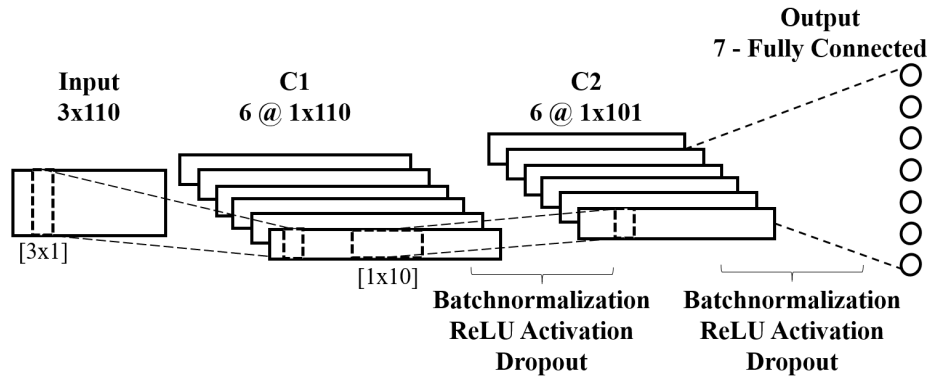
Hình 3.2: Biểu diễn FFT của người thứ 3 (subject 3)

3.2 Triển khai mô hình và đánh giá

3.2.1 Huấn luyện mô hình

Dữ liệu sau khi được tiền xử lý sẽ được đưa vào 3 mô hình để dự đoán nhãn đầu ra. Mô hình sử dụng CCA được huấn luyện bình thường, trong khi đó M-CNN và C-CNN được huấn luyện dạng User-Dependent như sau

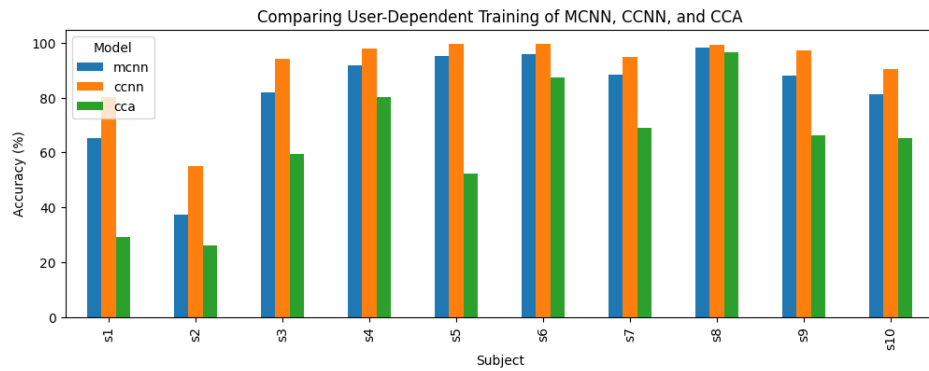
- User-Dependent Training (UD Training): Dữ liệu từ một subject được sử dụng để huấn luyện mô hình và được kiểm tra bằng dữ liệu của chính subject đó. Để thực hiện điều này, áp dụng kiểm tra chéo 10 lần (10-fold cross validation). Bộ dữ liệu được chia dựa trên quy trình kiểm tra chéo 10 lần trên các đoạn dài 1 giây của các epoch. Tổng số đoạn dài 1 giây trong tập huấn luyện là 648 và trong tập kiểm tra là 72. Các tham số cuối cùng của mạng được chọn là: $\alpha = 10^{-3}$, $momentum = 0.9$, $D = 0.25$, $L = 10^{-4}$, $E = 50$, $B = 64$.



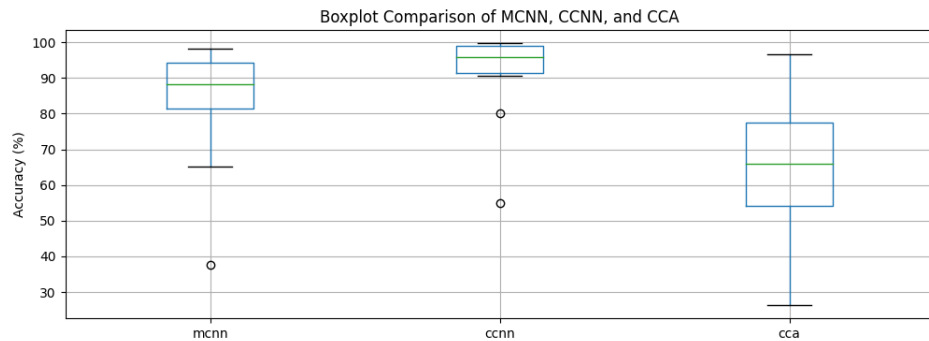
Hình 3.3: Minh hoạ kiến trúc CNN[2]

3.2.2 Kết quả và đánh giá

- Hình sau thể hiện các kết quả thu được của cả 3 phương pháp sử dụng trên bộ Dataset thuộc 10 subjects với cửa sổ $W = 1$. Có thể thấy rằng phương pháp UD C-CNN cho kết quả cao nhất với $90.82 \pm 13.19\%$, tiếp theo đó là UD M-CNN với $82.36 \pm 17.51\%$. Hai phương pháp sử dụng mạng CNN cho kết quả vượt trội của CCA với tỉ lệ chính xác là $62 \pm 21.5\%$



Hình 3.4: Tỷ lệ chính xác của các mô hình



Hình 3.5: Tính ổn định của các mô hình

- Hình ở trên thể hiện mức độ ổn định chính xác của các dự đoán trên các subjects. Giống với kết quả vừa nêu, C-CNN cho khả năng dự đoán chính xác ở mức rất cao ở tất cả subjects, còn M-CNN và CCA thì có mức độ ổn định khá.

Chương 4

Kết luận

Trong bài báo cáo đã thực hiện đề cập đến 3 phương pháp để phân loại tín hiệu EEG SSVEP. Phương thức CCA mặc dù thực hiện triển khai nhanh và tường minh tuy nhiên chưa đạt được độ chính xác quá cao trong việc phân loại. Hai phương pháp sử dụng mạng tích chập CNN với các tính chỉnh đầu vào được biến đổi sang miền tần số cho hiệu quả vượt trội so với phương pháp phân loại bằng CCA. Việc mô hình được huấn luyện theo dạng User-Dependent cho phép mô hình học được đặc trưng của tín hiệu subject nhanh hơn. Với một số thực nghiệm khác, các dạng UD cho hiệu suất vượt trội hơn User-Independent (UI) với cùng một mô hình. Tuy nhiên, việc cài đặt UD khó khăn hơn nhiều trong thực tế, trong khi UI có thể được thiết kế để hoạt động với bất kỳ subject riêng lẻ nào.

Các hệ thống SSVEP BCI trong thực tế có thể sử dụng mô hình CNN để thực hiện tác vụ dự đoán với thời gian hợp lý, thông qua đó cho phản hồi nhanh ở các ứng dụng thực tế. Trong tương lai, có thể áp dụng thêm nhiều mô hình học sâu khác để thực hiện tác vụ SSVEP và BCI liên quan.

Tài liệu tham khảo

References

- [1] Masaki Nakanishi et al. “A comparison study of canonical correlation analysis based methods for detecting Steady-State visual evoked potentials”. In: *PLoS ONE* 10.10 (Oct. 2015), e0140703. DOI: 10.1371/journal.pone.0140703. URL: <https://doi.org/10.1371/journal.pone.0140703>.
- [2] Aravind Ravi et al. “Comparing user-dependent and user-independent training of CNN for SSVEP BCI”. In: *Journal of Neural Engineering* 17.2 (Jan. 2020), p. 026028. DOI: 10.1088/1741-2552/ab6a67. URL: <https://doi.org/10.1088/1741-2552/ab6a67>.