ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10448073

TENSOR DECOMPOSITION-BASED DATA FUSION FOR BIOMARKER EXTRACTION FROM MULTIPLE EEG EXPERIMENTS

K.R. Stunnenberg, R.C. Hendriks*, J.L. Vroegop†, M.L. Adank† and B. Hunyadi**

∗ Signal Processing Systems group, Delft University of Technology, the Netherlands
† Dep. of Otorhinolaryngology and Head and Neck Surgery, Erasmus MC, Rotterdam, the Netherlands

ABSTRACT

The pursuit of sensitive and dependable biomarkers capable of capturing the neural processes associated with cognition is a prominent area of interest. Event-related potentials (ERPs) hold significant promise for assessing cognitive dysfunction in various neurological disorders. However, existing data analysis techniques often underutilize the available data and may benefit from potential enhancements. In this paper, we investigate biomarker extraction methods based on two ERP experiments. First, we derive average ERPs from the electroencephalography (EEG) recorded during each experiment and store them in third-order tensors with subjects, channels and time samples along the three modes. Then, we extract biomarkers from these datasets via tensor decompositions. We compare single tensor decompositions and joint tensor decompositions that fuse the data from the individual tensors. In a simulated ERP experiment we compare the benefits and limitations of different tensor-based data fusion methods. Finally, we investigate their performance on a real dataset obtained from schizophrenia patients.

Index Terms— tensor decompositions, data fusion, ERP, EEG, biomarker

1. INTRODUCTION

Identifying biomarkers could offer substantial benefits in various aspects and challenges that arise with a psychiatric disorder, such as early diagnosis or following disease progression. Event-related potentials (ERPs) derived from cognitive EEG experiments appear promising for extracting biomarkers. The classical way to analyse ERP data, i.e. comparing peak amplitudes and delays [1], is often expanded with machine-learning[2]. However, these approaches often focus on classification accuracy, but lack the insight into the differences between the healthy and the pathological signal. Decomposition techniques rely on the assumption that the observed ERP signals result from an additive mixture of various underlying factors or sources and that the strength of these sources varies per individual, and at least one of the sources shows a group difference between patients and healthy controls. EEG signals of multiple individuals (subjects) are inherently represented in a three-way tensor, with time, channel and subject modes. Therefore, tensor decompositions are the natural choice for their analysis [3].

Diagnostic procedures in clinical practice often involve multiple tests and experiments on patients. In addition to EEG measurements, this may include psychological tests, MRI data, or different types of EEG experiments. Such data can provide additional relevant information for the decomposition. Data fusion techniques combine different data sources and can exploit their complementary advantages [4]. Assuming the same patient-by-patient variability of underlying sources in both modalities, [5] proposed a joint tensor-matrix factorization model to extract detailed spatiotemporal information about the epileptic network from EEG and functional MRI data. However, the assumption of shared variability in all the sources may be too strong. Therefore, Acar et al. [6] proposed an advanced coupled matrix-tensor factorization (ACMTF) scheme that allows both shared and unshared factors in the data. Their method successfully revealed differences between patients with schizofrenia (SZ) and healthy controls (HC) based on functional MRI, structural MRI, and EEG data. Additionally, the assumption of perfectly aligned patient-by-patient variability in the shared factor may also be violated in real data. Therefore, [7] proposed a soft coupled decomposition for fusing EEG and fMRI data.

In this paper, we investigate the performance of these data fusion methods to classify groups of subjects with different ERP characteristics in a simulation study. Subsequently, we classify subjects with and without schizofrenia based on two EEG tensors recorded during two different auditory ERP experiments obtained from the study published in [1].

2. TENSOR DECOMPOSITIONS

A single data tensor can be decomposed by the canonical polyadic decomposition (CPD). The CPD decomposes an Nth order tensor into a linear combination of R terms, $\mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)}$. Each term in this formula is a rank-1 tensor, given by the outer product of each of the mode$-n$ signatures $\mathbf{u}_r^{(n)}$. The full tensor can be written as:

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \lambda_r \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)}, \tag{1}$$

where $\underline{\mathbf{X}} \in \mathcal{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is an Nth-order tensor and the λ_r's are the relative weights of the components. The CPD can be written in the following equivalent form:

$$\underline{\mathbf{X}} = \underline{\mathbf{\Lambda}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} \tag{2}$$

by forming a superdiagonal core tensor $\underline{\mathbf{\Lambda}} \in \mathcal{R}^{R \times R \times \cdots \times R}$ with the weights λ_r, and collecting the mode-n signatures into so-called factor matrices $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_R^{(n)}]$. The symbol $\times_n$ denotes the mode$-n$ product between a tensor and a matrix. For more information on tensor operations, see [8]. In the remainder of this paper we will use the following compact notation of the CPD with given weights and factor matrices: $[[\lambda; \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, ..., \mathbf{U}^{(N)}]]$. For example, in our case the ERP tensor $\underline{\mathbf{X}} \in \mathcal{R}^{\text{subjects} \times \text{time} \times \text{channels}}$ can be written in CPD form as $[[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]$, where the columns of the factor matrix $\mathbf{A}$ indicate the subject-by-subject strength of each component, the factor matrix $\mathbf{B}$ describes the channel distribution of the components, while $\mathbf{C}$ describes the component time courses. Note that real data is noisy, and therefore the CPD will be fitted to the data in the least-squares sense, i.e, we have an approximation instead of equality in (1) and (2).

In this paper, we consider datasets with two different experiments repeated per subject. It is expected that the two tensors with the ERP information from the two measurements have a correlation in the group differences. Therefore, instead of analyzing these tensors separately, it could be advantageous to use them simultaneously for finding biomarkers.

The most basic data fusion method assumes that the full subject-mode factor matrix is shared across the two datasets. Although originally proposed for fusing an EEG tensor and an fMRI matrix and hence called coupled matrix-tensor factorization (CMTF) [5], the cost function can simply be rewritten such that it applies to two tensors. For two ERP experiments with tensors $\underline{\mathbf{X}} \in \mathcal{R}^{N \times T \times Ch}$ and $\underline{\mathbf{Y}} \in \mathcal{R}^{N \times T \times Ch}$, with N, T and Ch are the number of subjects, time samples and EEG channels, respectively, this is given by

$$\begin{aligned} f(\lambda, \sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}) = ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||_F^2 \\ + ||\underline{\mathbf{Y}} - [[\sigma; \mathbf{A}, \mathbf{D}, \mathbf{E}]]||_F^2, \end{aligned} \tag{3}$$

where λ and σ are R-long vectors containing the relative weights of each component in tensor 1 and 2, $\mathbf{A} \in \mathcal{R}^{N \times R}, \mathbf{B} \in \mathcal{R}^{T \times R}, \mathbf{C} \in \mathcal{R}^{Ch \times R}, \mathbf{D} \in \mathcal{R}^{T \times R}$, and $\mathbf{E} \in \mathcal{R}^{Ch \times R}$ are the corresponding factor matrices.

Inspection of (3) shows that both datasets share the factor matrix $\mathbf{A}$. This means that if, for example, the first component in the decomposition tells something about a group difference, there is a corresponding ERP component from experiments 1 and 2 with that exact same group distribution. In the remainder of this paper, this method will be called fixed data fusion (FDF).

The fixed data fusion method is simple with relatively few parameters to tune. However, this method assumes that the coupled datasets only have shared components and may fail to capture the underlying patterns in the presence of both shared and unshared components. The hard constraint that the shared factors must be exactly the same in both decompositions may be too strict as well [7]. To ease the former restriction, structure revealing CMTF or advanced CMTF (ACMTF) was proposed [9], allowing for shared and unshared components to be present. The ACMTF model with R-components can be rewritten for an ERP tensor $\underline{\mathbf{X}}$ and an ERP tensor $\underline{\mathbf{Y}}$:

$$\begin{aligned} f(\lambda, \sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}) = ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||_F^2 \\ + ||\underline{\mathbf{Y}} - [[\sigma; \mathbf{A}, \mathbf{D}, \mathbf{E}]]||_F^2 + \beta||\lambda||_1 + \beta||\sigma||_1. \end{aligned} \tag{4}$$

The ℓ_1-regularization terms on the weight vectors enable the presence of unshared components across the two tensors. More specifically, these terms control the sparsity of the weight vectors. A zero entry for one tensor and a corresponding non-zero entry for the other tensor implements an unshared factor. In our case, the unshared components can still capture ERP components with a group difference that is heavily present in one dataset but not so strong in the other. However, a drawback of ACMTF is that it is sensitive to its parameters (number of components R and penalty parameter β) [7]. Advanced data fusion also relies on a hard coupling assumption, which means that the factor matrices are exactly the same across all datasets. To address this second restriction, a soft coupling approach can be used [7], where instead of assuming the exact same factor matrix across datasets and allowing unshared components, a relaxation on the shared factor matrices can be added. The cost function for a soft data fusion can be written as

$$\begin{aligned} f(\lambda, \sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{U}, \mathbf{D}, \mathbf{E}) = ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||_F^2 \\ + ||\underline{\mathbf{Y}} - [[\sigma; \mathbf{U}, \mathbf{D}, \mathbf{E}]]||_F^2 + \alpha||\mathbf{A} - \mathbf{U}||^2. \end{aligned} \tag{5}$$

Note that instead of the shared factor matrix $\mathbf{A}$, the second decomposition now uses a factor matrix $\mathbf{U}$. The similarity between these two factor matrices can be adjusted with the term $\alpha||\mathbf{A} - \mathbf{U}||_F^2$. For a large α value, the factor matrices $\mathbf{A}$ and $\mathbf{U}$ will be more similar, and for smaller values of α, the factor matrices can be different from each other.

3. SIMULATIONS

For simulations, we generated synthetic ERP data through BESA simulator software, which allows producing continuous EEG data that contain Event-Related Potentials.

We simulated ERP data based on the Mismatch Negativity (MMN) paradigm [10], involving ERPs in response to standard and deviant sounds. These waveforms are illustrated in Fig. 1(a). In healthy individuals, there is a difference in the ERP response of standard and deviant stimuli at the so-called N1 and P2 peaks (i.e. at the second positive and second negative peaks of the waveform). To evaluate the performance of the data fusion methods, we generated data for two groups of subjects exhibiting distinct responses to both the standard and

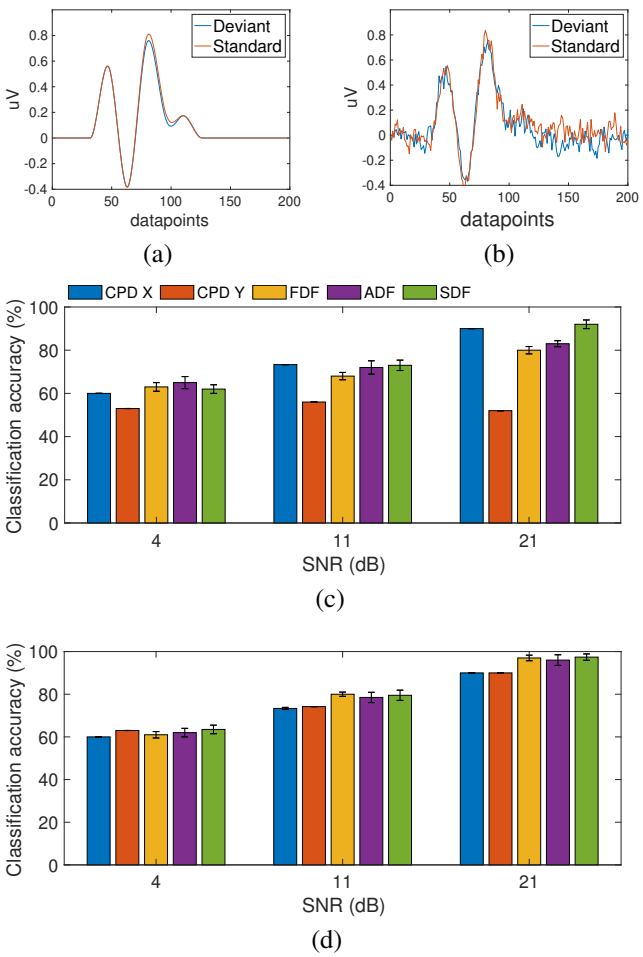

Fig. 1. Simulation results: (a) Clean waveform. (b) Noisy waveform (11 dB). (c) Classification results scenario 1. (d) Classification results scenario 2.

deviant stimuli within each group. We generated continuous 33-channel EEG data with 200 repetitions of ERPs embedded in it for 30 subjects. We generated data multiple times with different noise levels (4, 11 and 21 dB), including both alpha activity and white noise. Although the template ERP waveforms are identical within a group, there will be subject-specific differences in the ERP waveforms due to the additive noise. The continuous EEG with embedded ERPs were segmented into 200 samples around the simulated stimulus onset and averaged ERPs were created per subject. Examples of the averaged ERP waveforms from one subject can be seen in Fig. 1(b).

Two scenarios will be simulated, these are, (1) the group difference is only present in the deviant tensor, and (2) the group difference is present in both the deviant and standard tensors.

The amount of group difference (deviant P2 peak of 0.76uV in group 1 vs 0.85 in group 2; standard P2 peak 0.81 in group 1 and 0.93 in group 2) and noise levels were set such that the classification results range from "bad" to "good" to see the difference between the methods.

Using these waveforms, two tensors will be created: the deviant tensor $\underline{\mathbf{X}} \in \mathcal{R}^{N\times T\times Ch}$ and the standard tensor $\underline{\mathbf{Y}} \in \mathcal{R}^{N\times T\times Ch}$, where $N = 30$ is the number of subjects, $T = 200$ is the number of time samples and $Ch = 33$ is the number of channels. After the decompositions, the statistically significant components are selected with a two-sample t-test of the column of the subject factor matrix A (and U in case of SDF), and the classification of subjects is done with k-means clustering on the selected columns of the factor matrix. We tested different values of R between 1 and 6, and different values of the regularization parameters α and β, between 0.1 and 0.001.

Figs. 1(c) and 1(d) show classification accuracy for scenario 1 and scenario 2, respectively. For each method and scenario, the result obtained with the best combination of R and regularization parameter are shown. In most cases $R = 2$ was the best choice, except for ADF in the first scenario, which seemed to be very sensitive to the choice of β. As expected, classification results improve with higher SNR and for the first scenario, classification remains around chance level for the tensor $\underline{\mathbf{Y}}$. Moreover, if there is a difference present in both tensors (scenario 2), all data fusion methods outperform the single CPD especially for higher SNRs. On the contrary, when the second tensor does not have a group difference (scenario 1), FDF results end up in between the results of the two single CPDs for higher SNR values. For the lower SNR values, all methods perform rather poor. Compared to FDF, both ADF and SDF provide better results due to their flexibility.

4. SCHIZOPHRENIA DETECTION

We further evaluate the capabilities of the data fusion models for extracting schizofrenia biomarkers on a public dataset. Ford et al. [1] performed ERP experiments on 26 patients with SZ and 22 HC ranging from 19 to 61 years old. After the publication, additional measurements were released on Kaggle (`https://www.kaggle.com/datasets/broach/button-tone-sz`) for a total of 81 subjects (49 SZ and 32 HC), which is what we will use in our experiments. The released data is pre-processed, including filtering and artifact removaland baseline correction. The data we use was obtained using the following two experiments: (1) Subjects press a button and hear a 1000 Hz, 80 dB tone and (2) Subjects passively listen to a 1000 Hz, 80 dB tone.

Ford et al. [1] analyzed the ERP data to reveal differences in both groups. On average, they found slight differences between the N1 and P2 peaks between SZ and HC, however, these differences were not statistically significant. Statistically significant differences were found in the readyness potential (RP) preceding button press in experiment (1). In our experiments we investigate if these group differences could be used for classifying individual SZ and HC subjects.

The experiments involved $N = 81$ subjects, $T = 1700$ time samples per ERP and $E = 9$ electrodes. The data from experiment (1) were stacked into the tensor $\underline{\mathbf{X}} \in \mathcal{R}^{81\times1700\times9}$, while data from experiment (2) were stacked into $\underline{\mathbf{Y}} \in \mathcal{R}^{81\times1700\times9}$.

The tensor decomposition and classification strategy is similar to the simulation study in Sec. 3. The classification accuracy of the different methods can be seen in Fig. 2(a). The classification accuracy of the single tensor decompositions are 65% and 75%, respectively. The data fusion methods do not outperform the single CPD on $\underline{\mathbf{Y}}$, which suggests that the group differences in the two experiments are not shared. The overall result of 75% is not yet satisfactory in a clinical setting. We hypothesized that the group differences may be masked by differences in age, considering the fact that ERP waveforms change when a subject gets older. To test this hypothesis, we perfomed a second analysis where only subjects above 30 years of age are considered. As shown in Fig. 2(b), this increases classification accuracy to around 80% for the single CPD with $R = 2$ for experiment 2. However, the data fusion methods stay between the results of the single CPDs. From the two extracted components, one is statistically significant, which is shown in Fig. 2(c). The waveform seems to be related to the N1 peak, which suggests that the clinical difference related to this feature has differentiating power in the ERP, even though this difference was significant without CPD. The other, non-significant component was a combination of the N1 and P2 peaks. RP was not extracted by CPD.

For comparison, we performed k-means clustering on various combination of features including N1 and P2 peak amplitudes and delays from different electrodes. Classification accuracy stayed below 70% for all combinations, even when only the age > 30 group was considered.

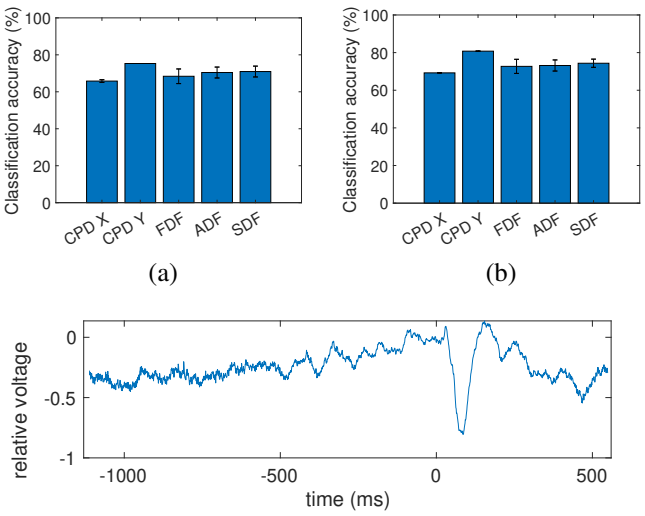

Fig. 2. Schizophrenia detection: (a) Results for all subjects. (b) Results for subjects with age > 30. (c) Waveform component 1 CPD $\underline{\mathbf{Y}}$ for subjects with age > 30.

5. DISCUSSION

Tensor decomposition-based data fusion methods combine information from multiple datasets by extracting shared factors that offer insight into the underlying neurological sources contributing to the differences in ERP responses. In the artificial simulations, we showed that data fusion methods can outperform the single decompositions if the additional dataset has shared differences between the two groups. If one of the two tensor does not have a group difference, fixed data fusion methods are more of a compromise between the single decomposition results. With optimal parameter settings, flexible data fusion methods outperform fixed data fusion; SDF being able to perform as well as the decomposition of the single tensor containing the group difference. Using data fusion methods opens up more possibilities and is not limited to combining ERP tensors, but can also be done with MRI data or even some demographic data, psychological score data [11], or movement score data.

Our results on the schizophrenia dataset show that by using tensor decompositions, the classification results are clearly much higher than only using amplitudes and delays. However, the data fusion methods do not perform better than the single decomposition of experiment two. Based on the simulations performed in Sec. 3, this can be explained: the data fusion classification results reach a value that is between the results of the single decompositions if the shared group difference is not present. This most likely means that the variation of the response between passively and actively listening to sound is too large.

6. CONCLUSION

In this paper we have utilized joint tensor decomposition-based data fusion techniques to combine multiple ERP datasets. We conclude that by using tensor decompositions, ERP data can be decomposed into different factors that can be used as biomarkers for classifying subjects. By performing simulations on artificial ERP data, we have shown that the different data fusion methods perform better on two ERP tensors compared to a single tensor decomposition. However, this is only the case if the group difference is shared among both datasets. When this condition is violated, fixed data fustion technique fall behind. Flexible fusion techniques can perform similarly to a single tensor decomposition in case optimal parameters are selected. However, this may be challenging in a real scenario. We have also shown on a real dataset that tensor decompositions not only have the potential to classify subjects, but also provide a better understanding about the electrophysiological signatures of the underlying neurological differences.

7. REFERENCES

[1] J. M. Ford, V. A. Palzes, B. J. Roach, and D. H. Mathalon, "Did I do that? abnormal predictive processes in schizophrenia when button pressing to deliver a tone," *Schizophrenia bulletin*, vol. 40, no. 4, pp. 804–812, 2014.

[2] Daniel Stahl, Andrew Pickles, Mayada Elsabbagh, Mark H Johnson, BASIS Team, et al., "Novel machine learning methods for erp analysis: a validation from research on infants at risk for autism," *Developmental neuropsychology*, vol. 37, no. 3, pp. 274–298, 2012.

[3] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of eeg signals: a brief review," *Journal of neuroscience methods*, vol. 248, pp. 59–69, 2015.

[4] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[5] B. Hunyadi, W. Van Paesschen, M. De Vos, and S. Van Huffel, "Fusion of electroencephalography and functional magnetic resonance imaging to explore epileptic network activity," in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 240–244.

[6] E. Acar, C. Schenker, Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data," *Frontiers in neuroscience*, vol. 13, pp. 416, 2019.

[7] C. Chatzichristos, S. Van Eyndhoven, E. Kofidis, and S. Van Huffel, "Coupled tensor decompositions for data fusion," in *Tensors for data processing*, pp. 341–370. Elsevier, 2022.

[8] A. Cichocki, N. Lee, I. Oseledets, A. Phan, Q. Zhao, D. P. Mandic, and et al., "Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions," *Foundations and Trends® in Machine Learning*, vol. 9, no. 4-5, pp. 249–429, 2016.

[9] E. Acar, E. E. Papalexakis, G. Gürdeniz, M. A Rasmussen, A. J. Lawaetz, M. Nilsson, and R. Bro, "Structure-revealing data fusion," *BMC bioinformatics*, vol. 15, no. 1, pp. 1–17, 2014.

[10] R. Naatanen, *Attention and brain function*, Hillsdale, NJ: Erlbaum, 1992.

[11] E. Kinney-Lang, A. Ebied, B. Auyeung, R. F. M. Chin, and J. Escudero, "Introducing the joint eeg-development inference (jedi) model: a multi-way, data fusion approach for estimating paediatric developmental scores via eeg," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 3, pp. 348–357, 2019.