Diagnosis of attention deficit and hyperactivity disorder (ADHD) using Hidden Markov Models

M.C. Maya-Piedrahita, D. Cárdenas-Peña, A.A. Orozco-Gutierrez
Automatics Research Group
Universidad Tecnológica de Pereira
Pereira, Risaralda, 660003
e-mail: {camila.maya,dcardenasp,aaog}@utp.edu.co

***Abstract*—Attention deficit hyperactivity disorder (ADHD), most often present in childhood, may persist in adult life, hampering personal development. However, ADHD diagnosis is a real challenge since it highly depends on the clinical observation of the patient, the parental and scholar information, and the specialist expertise. Despite demanded objective diagnosis aids from biosignals, the physiological biomarkers lack robustness and significance under the non-stationary and non-linear electroencephalographic dynamics. Therefore, this work presents a supported diagnosis methodology for ADHD from the dynamic characterization of EEG based on hidden Markov models (HMM) and probability product kernels (PPK). Relying on the impulsivity symptom, the proposed approach trains an HMM for each subject from EEG signals at failing rewarded inhibition tasks. Then, PPK measures the similarity between subjects through the inner product between their trained HMMs. Therefore, a support vector machine supports ADHD diagnosis as a classification task using PPK as the inner product operator. Results in a real EEG dataset evidence that the proposed approach achieves an 90.0% accuracy rate, outperforming log-likelihood features as baseline HMM-based features. Besides, achieving such an accuracy at the highest reward level supports that ADHD patients seem to be particularly sensitive to the reward presence when they execute specific tasks.**

Keywords–Attention deficit and hyperactiviy disorder; Hidden Markov models; Electroencephalography; Time-Series similarity

I. Introduction

As a neuropsychiatric disorder, Attention Deficit and Hyperactivity (ADHD) onsets on childhood inducing learning problems, low self-esteem, and substance use in adolescence [1]. Usually, ADHD diagnosis relies on information collected by family members and teachers and clinical observation. However, the analysis depends on the specialist expertise, leading to high misdiagnosis rates [2]. To find a more specific and sensitive diagnosis, strategies integrating biomarkers and regular scientific assessments are required [3].

In an attempt to describe reliable biomarkers for ADHD identification, current researches rely on the brain physiological response assessed through electroencephalography (EEG) to better understand neuropathologies [4]. For instance, the theta-beta power ratio (TBR) contrasts slow waves (4-7 Hz) and fast waves (13-30 Hz), under the assumption of more powerful slow waves in ADHD than in controls on resting state. Another proposed physiologically interpretable EEG feature corresponds to the P300 that expects a latency larger than the typical 300 ms after a stimulus is presented with lower amplitudes in ADHD subjects. Also, the error-related negativity (ERN) wave, evoked between 50-100 milliseconds at the frontal-central region when incorrect responses occur, has been considered to characterize ADHD children based on the impulsivity symptom [5]. Despite their physiological ground, recent works challenge the robustness and significance of TBR, P300, and ERN as ADHD biomarkers under the non-stationary and non-linear temporal dependencies of the EEG [6]–[8].

For coping with temporal dependencies, dynamic stochastic tools turn out to be more appropriate. Particularly, hidden Markov models (HMM) unravel temporal dynamics of time series by approximating the transition probability from a latent state to another, assuming a parametric distribution on the observed signal [9]. As an advantage, HMMs allow classifying signals with varying lengths by training an individual model per class and labeling a new signal according to the maximum a posteriori probability (MAP) criterion [10]. Another approach considers the Kullback-Leibler (KL) divergence between HMMs to feed distance-based classifiers [11]. However, KL divergence lacks a closed-form solution for HMMs, resulting in expensive Monte Carlo approximations [12]. Besides, the KL divergence does not satisfy the triangular inequality and symmetry properties required for the distance that parameterizes the classifier [13]. Dealing with these two last issues, the stationary cumulative distribution function of HMMs fulfills the distance properties [14]. Despite the cumulative functions can be directly compared over the observation space span, it is designed for unidimensional observations only [15].

Aiming to support ADHD diagnosis from EEG, this work proposes a classification methodology that assesses the similarity between time-series using the probability product kernel (PPK) between their corresponding HMM. In this regard, PPK gathers the affinity of all jointly observable sequences through the Bhattacharya distance between two HMMs [16]. Since introduced PPK behaves as an inner product in the space of probability distributions, it feeds a support vector machine that performs the classification of new EEG. We evaluate the proposed approach in the supported diagnosis of ADHD children, using EEG signals recorded while performing a cognitive paradigm magnifying the impulsivity symptom. The paradigm, termed Reward Stop-Signal Task (RSST), allows evaluating the classification results at three different rewards and two reward displaying orders. The comparison of the proposed PPK-based against the Log-likelihood-based features demonstrates that PPK improves the average classification accuracy while reducing the standard deviation.

The paper is organized as follows: Section 2 describes the proposed similarity between time series with PPK approach. Section 3 presents the experimental setup, explaining the considered EEG dataset, the HMM training, and parameter tuning. Section 4 focuses on the attained supported diagnosis results. Finally, Section 5 discusses the main findings and proposes a future work.

II. Similarity between time series

Let a set of N time series $\mathcal{X}=\{\boldsymbol{X}_i\in\mathbb{R}^{D\times T_i}\}_{i=1}^{N}$ where each $\boldsymbol{X}_i=\{\boldsymbol{x}_{i0},\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT_i}\}$ corresponds to an ordered sequence of T_i vector-valued observations $\boldsymbol{x}_{it}\in\mathbb{R}^D$ belonging to a D-dimensional Euclidean space. A single Hidden Markov Model (HMM) describes the stochastic behavior of each $\boldsymbol{X}_i$ by finding the sequence of hidden states $\boldsymbol{q}_i=\{q_{i0},q_{i1},\dots q_{iT_i}\}$ that maximizes the likelihood $p(\boldsymbol{X}_i|\boldsymbol{q}_i,\theta_i)$, with $q_{it}\in\{1,\dots,M\}$ as one of M possible states. θ_i comprises the following parameters of the i-th HMM:

- The initial state probability distribution $\boldsymbol{\pi}=[\pi_m=p(q_0=m)]$, $\boldsymbol{\pi}\in[0,1]^M$ with constraint $\sum_{m=1}^{M}\pi_m=1$.
- The state transition probability distribution $\boldsymbol{A}=[a_{mn}=p(q_t=m|q_{t-1}=n)]$, $\boldsymbol{A}\in[0,1]^{M\times M}$, constrained by $\sum_{m=1}^{M}a_{mn}=1$.
- The emission probability density function that is considered as Gaussian $p(\boldsymbol{x}_t|q_t=m)=\mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m)$, for $m\in\{1,\dots,M\}$.

To learn the parameter set θ_i, the Expectation-Maximization algorithm maximizes the likelihood in a two-step iterative way: The E-step estimates the most likely sequence of hidden states $\boldsymbol{q}_i$, and the M-step updates θ_i from the state posterior marginals.

Then, the similarity between a pair of time series $\boldsymbol{X}_i,\boldsymbol{X}_j$ can be assessed through the corresponding HMMs with parameters θ_i,θ_j using the probability product kernel between distributions (*PPK*) [16]:

$$\kappa(\boldsymbol{X}_i,\boldsymbol{X}_j)=\big\langle p(\boldsymbol{X}|\theta_i),p(\boldsymbol{X}|\theta_j)\big\rangle_{L_2}=\int p(\boldsymbol{X}|\theta_i)p(\boldsymbol{X}|\theta_j)d\boldsymbol{X} \quad (1)$$

where $\langle\cdot,\cdot\rangle_{L_2}$ stands for the inner product in the space of probability functions, and the integral accounts for all potential observable sequences.

Taking into account the factorization of the HMMs, Equation (1) can be approximated for a given sequence length T as in Algorithm 1, where $\psi_{mn}\in\mathbb{R}$ corresponds to the inner product between the observation pdfs from states m and n, which for two Gaussians $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m)$ and $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_n)$ is computed in a closed form as [17]:

$$\begin{aligned}
\psi_{mn}&=\langle p(\boldsymbol{x}|q_m),p(\boldsymbol{x}|q_n)\rangle_{L_2}\\
\psi_{mn}&=\langle\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m),\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_n)\rangle_{L_2}\\
\psi_{mn}&=(2\pi)^{(1-2\rho)D/2}\rho^{-D/2}\left|\boldsymbol{\Sigma}^{\dagger}\right|^{1/2}|\boldsymbol{\Sigma}_m|^{-\rho/2}|\boldsymbol{\Sigma}_n|^{-\rho/2} \quad (2)\\
&\quad\times\exp\left(\frac{\rho}{2}\left(\boldsymbol{\mu}_m^{\top}\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m+\boldsymbol{\mu}_n^{\top}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n-\boldsymbol{\mu}^{\dagger\top}\boldsymbol{\Sigma}^{\dagger}\boldsymbol{\mu}^{\dagger}\right)\right)\\
\boldsymbol{\Sigma}^{\dagger}&=\left(\boldsymbol{\Sigma}_m^{-1}+\boldsymbol{\Sigma}_n^{-1}\right)^{-1}\\
\boldsymbol{\mu}^{\dagger}&=\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m+\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n
\end{aligned}$$

where $\boldsymbol{\Sigma}^{\dagger}$ stands for the pseudo-inverse of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ correspond the mean and covariance of the n-th Gaussian pdf, respectively.

Algorithm 1 Probability Product Kernel for HMM with Gaussian emissions

Input: HMM parameters $\boldsymbol{\pi}^i,\boldsymbol{A}^i,\boldsymbol{\mu}_m^i,\boldsymbol{\Sigma}_m^i,\boldsymbol{\pi}^j,\boldsymbol{A}^j,\boldsymbol{\mu}_n^j,\boldsymbol{\Sigma}_n^j$ and PPK parameter T

Output: PPK approximation $k_{ij}\approx\kappa(\boldsymbol{X}_i,\boldsymbol{X}_j)$ between time series $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$

$\varphi_{mn}^0=\left(\pi_m^i\pi_n^j\right)^{1/2}\quad\forall n,m\in\{1,\dots,M\}$

for $t=1\dots T$ **do**

$\varphi_{mn}^t=\sum_{m'=1}^{M}\sum_{n'=1}^{M}\psi_{m'n'}\left(a_{mm'}^i a_{nn'}^j\right)^{1/2}\varphi_{m'n'}^{t-1}\quad\forall n,m\in\{1,\dots,M\}$

$k_{ij}=\sum_{m=1}^{M}\sum_{n=1}^{M}\psi_{mn}\varphi_{mn}^T$

Applying PPK at all HMMs pairs in $\mathcal{X}$ yields a symmetric definite positive kernel matrix $\boldsymbol{K}_{\mathcal{X}}(M,T)\in\mathbb{R}^{N\times N}$ with elements $k_{ij}(M,T)$. Note that the similarity measure between time series through their describing HMMs depends on two free parameters, namely, the number of discrete hidden states M considered to learn the HMM, and the sequence length T up to which the models are compared.

III. Experimental setup

A. Dataset and preprocessing

To evaluate the proposed time series classification methodology, we consider a dataset of electroencephalography (EEG) recordings from 67 children labeled as either healthy control or attention deficit hyperactivity disorder (ADHD) patient. EEG signals were recorded using the *Reward Stop-Signal Task* (RSST) paradigm, where participants must press a key each time they face a frequent stimulus called *Go*, unless a rare stimulus called *Stop* appears after the *Go*. Each time the *Go* is presented defines a trial. Participants executed the paradigm during four blocks of four minutes and about 300 trials each, receiving a reward if they successfully inhibit from pressing the key. The paradigm includes a *Smiley* sticker, a *Low* amount of candies, and a *High* amount of candies rewards. If the reward decreases from block to block, the participant belongs to the decreasing condition group (DC). On the contrary, participants belong to increasing condition group (IC) when the paradigm starts with a *Low* and ends with a *High* reward. Only trials where the subjects pressed a key despite the *Stop* stimulus (failed inhibitions) compose the time series dataset, due to such cognitive response has been identified as a biomarker of neurological disorders [18], [19]. Table I summarizes the number of subjects per class and condition, along with the number of failed inhibitions of each subject at each reward.

Condition	Class	N	*Smiley*	*Low*	*High*
Decreasing	Control	17	32.4±11.3	29.8±7.6	62.8±17.6
	ADHD	13	31.4±9.1	40.4±13.1	77.4±25.7
Increasing	Control	20	34.2±11.4	75.1±18.5	33.7±12.0
	ADHD	17	35.8±15.3	72.5±29.3	33.4±13.8

TABLE I: Number of subjects and average number of failed inhibitions in the dataset.

Regarding the time series details, EEG signals were recorded at 250 Hz and 32 channels distributed over the scalp, registering activity over the medial frontal [20], left frontal [19], ventromedial orbitofrontal, and prefrontal cortices known to evoke event-related negative (ERN) waves [21]. Firstly, each trial is trimmed 200 ms before and 800 ms after the *Go* stimulus, producing sequences of T_i=250 time instants. Then, we centered and normalized each channel regarding its mean and standard deviation on the pre-*Go* period as in Equation (3), where $\boldsymbol{E}_{ref}\{\cdot\}$ is the averaging operator over the time before the *Go*.

$$\hat{\boldsymbol{x}}_{it} = \frac{\boldsymbol{x}_{it} - \boldsymbol{E}_{ref}\{\boldsymbol{x}_{it}\}}{\sqrt{\boldsymbol{E}_{ref}\left\{(\boldsymbol{x}_{it} - \boldsymbol{E}_{ref}\{\boldsymbol{x}_{it}\})^2\right\}}} \tag{3}$$

B. HMM training

Since the RSST paradigm considers *Smiley*, *Low*, and *High* rewards, we train a single HMM for each of them, yielding three models per subject. In order to avoid overfitting, Expectation-Maximization (EM) algorithm learns HMM parameters using 70% of the failed inhibitions, and the remaining 30% stops the training. Figure 1, exemplifying the EM learning, evidences that the log-likelihood monotonically increases for the training EEG. On the contrary, the test data curve reaches a maximum at a few iterations, from which the model overfits the training data. Therefore, the hold-out appearance in HMM training not only guarantees the generalization of subject data but also reduces the number of EM iterations.

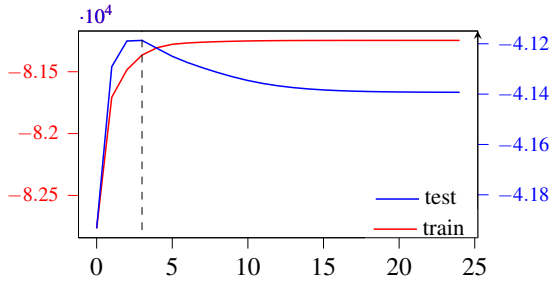
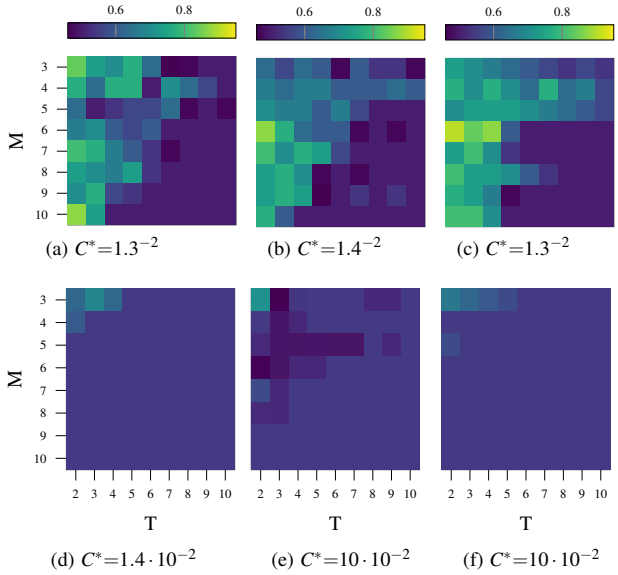


Figure 1: Log-likelihood along the EM iterations in the HMM training process for a control subject at *High* reward and M=3 states.

C. Parameter tuning and Classification

Given the set of trained HMMs, a support vector machine (SVM) fed by the probability product kernel carries out the classification of subjects into control or ADHD. Taking into account that the proposed methodology depends on the number of states, the sequence length, and the SVM box constraint, we tune such parameters using a 5-fold cross-validated grid search within $M\in\{3,\ldots,10\}$, $T\in\{2,\cdots,10\}$, and $C\in\{10^{-3},\cdots,10^{2}\}$. Figure 2 illustrates the attained classification accuracy averaged over the five test sets along the $M-T$ grid at the optimal box constraint for each reward and condition. For the parameter tuning, the proposed approach achieves the best performance at M=6, T=2, and *High* with 90.0% accuracy, and at M=3, T=2, and *Low* with 71.0% for DC and IC children, respectively. Note that for both conditions, the larger the sequence length, the worse the accuracy because all the kernel values decrease, difficulting the classifier training. On the other hand, increasing the number of states reduces the accuracy at IC, while reaching a maximum performance at DC. Hence, gaining HMM complexity benefits the decreasing condition and hampers the increasing one.



(a) C^*=1.3^{-2} (b) C^*=1.4^{-2} (c) C^*=1.3^{-2}

(d) C^*=1.4·10^{-2} (e) C^*=10·10^{-2} (f) C^*=10·10^{-2}

Figure 2: PPK classification accuracy for each evaluated parameter set at decreasing (top) and increasing (bottom) conditions. Accuracy is averaged over five test folds. Left to right: *Smiley*, *Low*, and *High* rewards.

IV. Results

Since the proposed methodology matches the time series in terms of their HMM, we compare against the classical HMM-based classification scenario, where a single HMM models each class and reward [10]. Then, an SVM classifies a target subject using as features the log-likelihoods (LL) of the subject trials given the HMM of their corresponding reward. We note that SVM parameters for LL were trained in the same PPK training scheme, using a Gaussian kernel.

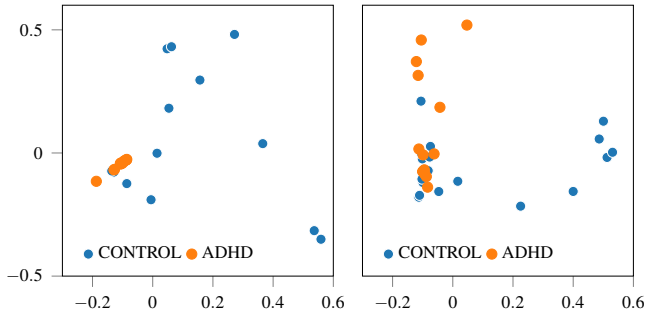
Table II summarizes the average accuracy with standard deviation attained by PPK and LL for considered conditions and rewards. Results evidence that the PPK approach outperforms LL at all rewards and conditions. Further, the more significant accuracies on DC than on IC prove that such a condition is more adequated for supported diagnosis purposes. Regarding the reward-condition pair, *High* at DC attains the best performance with a suitable standard deviation, which agrees with the RSST paradigm designed to highlight differences between ADHD and controls. As a result, the introduced PPK enhances the supported diagnosis of ADHD by assessing the similarity between the electrical activity of subjects.

For the sake of visualizing the separability between ADHD and control subjects, we applied Kernel Principal Component

Condition	*Approach*	*Reward*		
		Smiley	*Low*	*High*
DC	*LL*	73.3 ± 9	68.0 ± 7	64.4 ± 11
	PPK	**87.0 ± 13**	**87.0 ± 7**	**90.0 ± 8**
IC	*LL*	68.1 ± 17	62.8 ± 21	67.8 ± 20
	PPK	68.0 ± 12	71.0 ± 8	65.0 ± 8

TABLE II: Classification results for PPK and LL per conditions and rewards.

Analysis (KPCA) on the kernel matrix resulting from PPK. Figure 3 illustrates the data projected over the first two KPCA basis with the optimal parameters at the best performing reward for each condition. Data distribution proves that classes are more cluttered in IC than in DC so that PPK better defines a boundary between both classes in the latter condition. Therefore, PPK estimates similarities allowing gathering subjects according to their given diagnosis.



(a) DC *High* (b) IC *Low*

Figure 3: KPCA-based dataset projection for both conditions at the optimal parameters.

V. Discussion and Concluding Remarks

This paper proposes an ADHD supported-diagnosis methodology from EEG recording. The proposed methodology assesses the similarity between EEG signals using the PPK between their corresponding HMMs. Thanks to the HMM factorization, PPK can be approximated in a closed form for a given sequence length and number of hidden states. Therefore, the resulting kernel can feed classifiers based on inner-products, such as a support vector machine.

To test the methodology, we consider EEG signals recorded under the RSST paradigm at three reward levels and two ways to present rewards, termed conditions, yielding six EEG subsets. Since PPK depends on the number of states and sequence length, a cross-validation scheme searches for the highest accuracy in a parameter grid. Tuning results in Figure 2 shows that performance decreases as the sequences become longer. Further, either a few or a large number of states hinders the differences among subjects, so that there exist an intermediate M unraveling class differences. Consequently, optimal PPK parameters result in a complex enough measure enhancing the classification performance.

In order to compare the proposal, we consider class-wise and reward-wise LL as conventional features based on HMM for discriminating classes. Comparison results in Table II show that PPK outperforms LL at the decreasing condition while behaving similarly at the increasing one. Besides, since PPK deviations are shorter than LL, the proposed approach yields a more confident classification performance. Those attained accuracies agree with the resulting subject distribution illustrated in the KPCA projection (Figure 3), where an SVM can design a straightforward boundary in the DC subjects. Therefore, PPK develops a similarity measure that identifies the class differences supporting the ADHD diagnosis.

Regarding the RSST paradigm, we only considered failed inhibitions trials relying on the executive dysfunction and the deregulation of the inhibitory functions due to ADHD. From attained accuracies, we identified *High* as the best reward for DC condition, agreeing with the motivational changes associated with the attention deficit [22]. Finally, the decreasing condition generally improves the diagnosis in comparison with increasing one. Such a fact can be related to a more efficient inhibition modulation when starting with higher valued rewards, which has been previously reported for rewarded paradigms [23].

As a future work, we will develop a cost function measuring the agreement between the probability product kernel and objective labels, so that the number of states and sequence length can be optimized in a scheme more efficient than a grid search. Additionally, we plan to evaluate the performance of the generalized PPK metric that introduces a power constant allowing to weight events according its occurrence probability. Lastly, we aim at introducing spectral information by decomposing the signal into brain rhythms so that the methodology automatically identifies the most relevant ones for the task at hand.

VI. Acknowledgment

This work was supported by Vice-rectory for research from Universidad Tecnológica de Pereira and the research project number 111080763051 funded by MinCiencias. Authors also thank the Master's program in Electrical Engineering from UTP.

References

[1] S. Khoshnoud, M. A. Nazari, and M. Shamsi, "Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals," *Journal of integrative neuroscience*, 2018.

[2] V. del Barrio, "Diagnostic and Statistical Manual of Mental Disorders," in *Encyclopedia of Applied Psychology, Three-Volume Set*, 2004.

[3] S. M. Snyder, T. A. Rugino, M. Hornig, and M. A. Stein, "Integration of an EEG biomarker with a clinician's ADHD evaluation," *Brain and Behavior*, 2015.

[4] M. Ahmadlou and H. Adeli, "Wavelet-synchronization methodology: A new approach for EEG-based diagnosis of ADHD," *Clinical EEG and Neuroscience*, 2010.

[5] L. Marquardt, H. Eichele, A. J. Lundervold, J. Haavik, and T. Eichele, "Event-related-potential (ERP) correlates of performance monitoring in adults with attention-deficit hyperactivity disorder (ADHD)," *Frontiers in psychology*, vol. 9, p. 485, 2018.

[6] J. F. Saad, M. R. Kohn, S. Clarke, J. Lagopoulos, and D. F. Hermens, "Is the Theta/Beta EEG Marker for ADHD Inherently Flawed?" *Journal of Attention Disorders*, vol. 22, no. 9, pp. 815–826, 2018.

[7] M. J. Groom, G. Scerif, P. F. Liddle, M. J. Batty, E. B. Liddle, K. L. Roberts, J. D. Cahill, M. Liotti, and C. Hollis, "Effects of motivation and medication on electrophysiological markers of response inhibition in children with attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 67, no. 7, pp. 624–631, 2010.

[8] S. Whitmont, R. Meares, E. Gordon, I. Lazzaro, and S. Clarke, "The Modulation of Late Component Event Related Potentials by Pre-Stimulus EEG Theta Activity in ADHD," *International Journal of Neuroscience*, vol. 107, no. 3-4, pp. 247–264, 2008.

[9] J. Tornero Lucas, "Machine Learning: Modelos Ocultos de Markov (HMM) y Redes Neuronales Artificiales (ANN)," p. 53, 2017. [Online]. Available: http://diposit.ub.edu/dspace/bitstream/2445/122446/2/memoria.pdf

[10] S. Watanabe, A. Nakamura, and B. H. Juang, "Structural bayesian linear regression for hidden markov models," *Journal of Signal Processing Systems*, vol. 74, no. 3, pp. 341–358, 2014.

[11] B.-H. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden markov models," *AT&T technical journal*, vol. 64, no. 2, pp. 391–408, 1985.

[12] C. Lu, J. M. Schwier, R. M. Craven, L. Yu, R. R. Brooks, and C. Griffin, "A normalized statistical metric space for hidden Markov models," in *IEEE Transactions on Cybernetics*, 2013.

[13] M. Falkhausen, H. Reininger, and D. Wolf, "Calculation of distance measures between hidden Markov models." *Simulation*, 1995.

[14] J. Zeng, J. Duan, and C. Wu, "A new distance measure for hidden Markov models," *Expert Systems with Applications*, 2010.

[15] E. Epaillard and N. Bouguila, "Data-free metrics for dirichlet and generalized dirichlet mixture-based hmms – a practical study," *Pattern Recognition*, vol. 85, pp. 207 – 219, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031830311X

[16] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *Journal of Machine Learning Research*, 2004.

[17] T. Jebara and R. Kondor, "Bhattacharyya and expected likelihood kernels," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2003.

[18] Y. Groen, A. A. Wijers, L. J. Mulder, B. Waggeveld, R. B. Minderaa, and M. Althaus, "Error and feedback processing in children with adhd and children with autistic spectrum disorder: an eeg event-related potential study," *Clinical neurophysiology*, vol. 119, no. 11, pp. 2476–2493, 2008.

[19] C. S. van Meel, D. J. Heslenfeld, J. Oosterlaan, and J. A. Sergeant, "Adaptive control deficits in attention-deficit/hyperactivity disorder (adhd): the role of error processing," *Psychiatry research*, vol. 151, no. 3, pp. 211–220, 2007.

[20] K. R. Ridderinkhof, M. Ullsperger, E. A. Crone, and S. Nieuwenhuis, "The role of the medial frontal cortex in cognitive control," *science*, vol. 306, no. 5695, pp. 443–447, 2004.

[21] H. Garavan, T. Ross, K. Murphy, R. Roche, and E. Stein, "Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction," *Neuroimage*, vol. 17, no. 4, pp. 1820–1829, 2002.

[22] A. C. Kelly, A. Scheres, E. S. SONUGA-BARKE, and F. X. CASTELLANOS, "11 functional neuroimaging of reward and motivational pathways in adhd," *Handbook of attention deficit hyperactivity disorder*, p. 209, 2007.

[23] P. M. Herrera, A. V. Van Meerbeke, M. Speranza, C. L. Cabra, M. Bonilla, M. Canu, and T. A. Bekinschtein, "Expectation of reward differentially modulates executive inhibition," *BMC psychology*, vol. 7, no. 1, pp. 1–10, 2019.