Efficient and Accurate Neural Fingerprints Obtained via Mean Curve Length of High Dimensional Model Representation of EEG Signals

Evrim Korkmaz Özay
Software Engineering Department
Beykent University
İstanbul, Turkey
evrimozay@beykent.edu.tr

Tolga Esat Özkurt
Department of Health Informatics
Middle East Technical University
Ankara, Turkey
ozkurt@metu.edu.tr

***Abstract*—In this study, we propose and evaluate a feature extraction methodology for the purpose of EEG-based person recognition. To this end, the mean curve length (MCL) was employed subsequent to the representation of EEG signals in an orthogonal geometry through High Dimensional Model Representation (HDMR). To analyze the effectiveness of the methodology, we executed it on a standard publicly available EEG dataset containing 109 subjects and acquired from 64 channels for eyes-open (EO) and eyes-closed (EC) resting-state conditions. The proposed feature was evaluated by comparing it to MCL, beta, and gamma band activities. According to the performance results, applying MCL to the output of the HDMR instead of raw data provides superior performances for identification and authentication. The attained results promise a novel simple, fast, and accurate biometric recognition scheme, named HDMRMCL.**

***Index Terms*—authentication, biometrics, EEG, HDMR, identification, mean curve length, resting-state**

I. Introduction

The upsurge of portable and wireless EEG systems allows for effective biometrics from brain signals [1]. A practical EEG-based biometrics requires computational efficiency as well as adequate recognition accuracy for reliability.

So far, most of the current literature apropos of EEG-based biometrics has typically relied on spectral measures (band power values) and functional connectivity metrics such as coherence and phase-locking value. According to a recent study that relied on neural sources extracted from magnetoencephalographic data [2], beta-band activity was found to yield the highest differentiation, while higher-frequency gamma band activity was the most robust for short data segments. Though less common, other types of features, particularly based on signal complexity, have also been attempted in EEG-based biometrics such as entropy measures [3]–[5], fractal dimension [6], [7] and the aperiodic component [8].

A recent study co-authored and conceptualized by one of us [9] showed that a very simple metric, namely "mean curve length" (MCL) is dependent upon only a summation of discrete-derivative of the crude sensor-level EEG signals, and it enables rapid recognition performances even when utilized with Euclidean distance for classification.

High Dimensional Model Representation (HDMR) is a divide-and-conquer algorithm proposed by Sobol [10] to express multivariate functions with lower-variate functions. HDMR is an efficient tool to tackle the curse of dimensionality in different multivariate problems. Recent studies have exploited HDMR for various reasons, such as reliability analysis for the estimation of the failure probability of a dynamic problem and to handle a geotechnical engineering problem [11], [12]. HDMR was also employed as a meta-modeling tool for multidimensional input-output systems [13], [14]. Recently, HDMR has been used on tensor-type data sets for high dimensional image processing [15], [16].

In this study, we propose HDMR as an EEG preprocessing tool in order to improve the performance of the brainwave-based biometric system suggested by [9]. Henceforth, one of HDMR's components was given as an input to MCL for distinctive feature extraction (HDMRMCL). Subsequently, we compared the recognition performance obtained with HDMRMCL to those obtained with MCL applied to low-pass filtered EEG resting-state data and the spectral activities within the frequency ranges of beta and gamma bands. In addition, we endeavored to investigate the sensitivity and performance of our proposed scheme with respect to data length, which is one of the considerable factors for the fulfillment of a practical real-time EEG-based biometrics system. By making all these investigations, HDMR provides a very accurate and rapid contribution to the biometric identification problem. To the best of our knowledge, this is the first study to utilize HDMR to preprocess EEG data.

This paper is organized as follows: Section II gives detailed information on the EEG data and the methods used for our proposed feature extraction scheme, HDMR, and MCL, respectively. This section also provides a brief explanation of the classification procedure performed for the identification and authentication of neural signals. We analyze and discuss the experimental results of the proposed EEG-based biometric identification system in Section III. Finally, conclusive remarks and potential future work are noted in Section IV.

II. Materials and methods

A. Dataset and preprocessing

We use PhysioNet EEG Motor Movement/Imagery Dataset that has frequently been employed in EEG-based biometric recognition [17]–[19]. This dataset was acquired from a BCI2000 system with 64 electrodes (http://www.bci2000.org). The sampling rate of the recorded signals is 160 Hz. The dataset contains EEG recordings of eyes-open (EO) and eyes-closed (EC) resting states of one-minute length from 109 participants. A low-pass finite impulse response (FIR) filter was applied to each segment with a cut-off frequency of 50 Hz. In this study, we split resting-state data into a number of segments with a fixed length varying from 0.1 s (16 data points and 600 segments) to 15 s (2400 data points and 4 segments) and take one of them arbitrarily as a test segment in a cross-validation manner.

B. Feature extraction

Our main purpose is to design a computationally efficient EEG-based biometric identification and authentication system with high recognition accuracy by utilizing HDMR and MCL collaboratively. As a feature extractor, MCL's simplicity provides low computational cost and high accuracy in a biometric identification system, while HDMR highlights the individual statistical characteristics of high-dimensional data. Thereby, we hypothesized that combining HDMR and MCL may help reveal superior subject-discriminative traits of individual EEG signals while retaining computational efficiency. Figure 1 displays the flow of the proposed EEG-based biometric identification system and exhibits how MCL is applied to the HDMR's three-way (trivariate) component in this study. Section II-B1 gives a brief explanation of HDMR on multi-dimensional datasets.

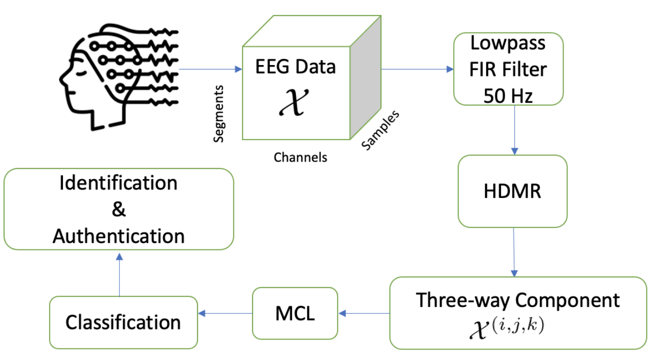

Fig. 1: The proposed EEG-based biometric identification system

1) High Dimensional Model Representation (HDMR): HDMR is a mathematical model that represents N-dimensional data (or tensor) with the sum of arrays in ascending dimensionality. These arrays are the orthogonal components of the model, based on projections of the data in Hilbert Space [20]. A certain element of an N-dimensional dataset can be evaluated with HDMR expansion as follows.

$$\begin{aligned} \mathcal{X}_{i_1 \ldots i_N} &= \mathcal{X}^{(0)} + \sum_{j_1=1}^{N} \mathcal{X}_{i_{j_1}}^{(j_1)} + \sum_{\substack{j_1,j_2=1 \\ j_1<j_2}}^{N} \mathcal{X}_{i_{j_1},i_{j_2}}^{(j_1 j_2)} \\ &+ \cdots + \mathcal{X}_{i_{j_1} i_{j_2} \ldots i_{j_N}}^{(j_1 j_2 \ldots j_N)}, \\ i_j &= 1,2,\ldots,n_j, \quad j=1,2,\ldots,N \end{aligned} \tag{1}$$

where $i_1 \ldots i_N$ are the element indices of the N-dimensional tensor, $\mathcal{X}$ (for this study it is the filtered dataset organized as 3-dimensional with segments, channels, and the number of samples for each person). The superscript indices at the right-hand side of Equation 1 denote the directions, while the subscripts indicate the elements of the HDMR components.

The right-hand side components of Eq. 1 stand for the tensors of the increasing number of dimensions. For example, $\mathcal{X}^{(0)}$ is a constant component of HDMR, $\mathcal{X}^{(j_1)}$s are one-way tensors (or vectors), and $\mathcal{X}^{(j_1 j_2)}$s are two-way tensors (or matrices), and so on. Weight vectors are defined under some conditions to determine HDMR components on the right-hand side. The first condition is the normalization condition over the weight vectors. For this work, we imposed the weights as coming from the uniform distribution for simplicity, as an example j^{th} direction's weight vector is $\frac{1}{n_j}$. The second condition is the vanishing condition imposed to obtain HDMR components uniquely. HDMR components are orthogonal to each other through the inner product.

$$\begin{aligned} &\sum_{i_j=1}^{n_j} W_{i_j}^{(j)} = 1, \quad \sum_{i_{j_l}=1}^{n_{j_l}} W_{i_{j_l}}^{(j_l)} \mathcal{X}_{i_{j_1} \ldots i_{j_k}}^{(j_1 \ldots j_k)} = 0, \\ &l = 1,2,\ldots,k, \quad k=1,2,\ldots,N, \quad j=1,2,\ldots,N \end{aligned} \tag{2}$$

Under the above conditions, the first component of HDMR to compute is the constant component, namely $\mathcal{X}^{(0)}$, and can be attained by projecting the dataset onto the weighted mean.

$$\mathcal{X}^{(0)} = \sum_{i_1=1}^{n_1} \cdots \sum_{i_N=1}^{n_N} \left[\prod_{k=1}^{N} W_{i_k}^{(k)} \right] \mathcal{X}_{i_1 \ldots i_N} \tag{3}$$

One-way components, $\mathcal{X}^{(j)}$ are determined with the exclusion of the j^{th} direction. This way the data are projected onto the j^{th} subspace.

$$\begin{aligned} \mathcal{X}_{i_j}^{(j)} &= \sum_{i_1=1}^{n_1} \cdots \sum_{i_{j-1}=1}^{n_{j-1}} \sum_{i_{j+1}=1}^{n_{j+1}} \cdots \sum_{i_N=1}^{n_N} \left[\prod_{\substack{k=1 \\ j \neq k}}^{N} W_{i_k}^{(k)} \right] \\ &\times \mathcal{X}_{i_1 \ldots i_N} - \mathcal{X}^{(0)} \\ &i_j = 1,2,\ldots,n_j, \quad j=1,2,\ldots,N \end{aligned} \tag{4}$$

Two-way components are computed by excluding two related directions of the two-dimensional subspace in the same manner.

$$\mathcal{X}^{(j,k)}_{i_j,i_k} = \sum_{i_1=1}^{n_1} \cdots \sum_{i_{j-1}=1}^{n_{j-1}} \sum_{i_{j+1}=1}^{n_{j+1}} \cdots \sum_{i_{k-1}=1}^{n_{k-1}} \sum_{i_{k+1}=1}^{n_{k+1}} \cdots$$
$$\cdots \sum_{i_N=1}^{n_N} \left[\prod_{\substack{k=1 \\ j\neq k}}^{N} W^{(k)}_{i_k} \right] \mathcal{X}_{i_1 \ldots i_N} - \mathcal{X}^{(k)}_{i_k} - \mathcal{X}^{(j)}_{i_j} - \mathcal{X}^{(0)}$$
$$j,k = 1,2,\ldots,N, \quad i_j = 1,2,\ldots,n_j,$$
$$i_k = 1,2,\ldots,n_k \tag{5}$$

Higher order HDMR components are determined similarly and the last term $\mathcal{X}^{(j_1 j_2 \ldots j_N)}$ is the residual after all HDMR components were subtracted from the data. Each one-way component of HDMR denotes the effect of the way (or variable) related to the component. Two-way and higher-order components give the correlated effect of the related ways. To utilize this property of HDMR on EEG-based biometric identification, first, we re-organize the EEG data as a three-dimensional tensor. Hence, for each participant, signals are organized as a 3D tensor $(\mathcal{X})$. The dimensions of the tensor are the number of segments, the number of channels, and the number of samples in each segment, respectively. The three-way component of HDMR is extracted from the expansion with the following exclusion step.

$$\begin{aligned} \mathcal{X}^{(1,2,3)}_{i,j,k} &= \mathcal{X}_{i,j,k} - \mathcal{X}^{(1,2)}_{i,j} - \mathcal{X}^{(1,3)}_{i,k} - \mathcal{X}^{(2,3)}_{j,k} \\ &- \mathcal{X}^{(1)}_{i} - \mathcal{X}^{(2)}_{j} - \mathcal{X}^{(3)}_{k} - \mathcal{X}^{(0)} \\ i &= 1,2,\ldots,n_1, j = 1,2,\ldots,n_2, \\ k &= 1,2,\ldots,n_3 \end{aligned} \tag{6}$$

The above component $(\mathcal{X}^{(1,2,3)}_{(i,j,k)})$ is the element of the residual term of HDMR located on i,j,k indices. This three-way component is obtained by the exclusion of the lower-order terms and in this way feature characteristics of the signals are purified and highlighted by HDMR. After the computation of HDMR's three-way component, MCL is applied to the purified signal as HDMR's three-way component.

2) Mean Curve Length (MCL): MCL of a time-series is the average of the absolute values of finite differences on this time-series [21]. It was used for epileptic seizure prediction successfully [22]–[24]. Despite being a very simple feature, MCL was recently shown to be highly discriminative for EEG biometrical recognition [9]. MCL for a signal with the length of N can be computed as follows

$$MCL = \frac{1}{N-1} \sum_{n=0}^{N-2} |x(n+1) - x(n)| \tag{7}$$

where $n = 1, 2, \cdots, N-1$.

C. Classification

We investigated the performance of the proposed feature extraction method HDMRMCL by setting up the very same classification method used in the recent MCL brainwave biometrics study in order to make a fair comparison [9]. The effect of HDMRMCL feature extraction scheme on the identification and authentication performances is evaluated with Mahalanobis distance-based nearest centroid classifier. The Classification was carried out with cross-validation by leaving out one segment for the test stage and using the other segments for the training stage. Both identification and authentication results were compared with those obtained with features via only MCL, beta (13-30 Hz; BETA), and gamma band (30-50 Hz; GAMMA) activities. The last two spectral bands were specifically chosen as they were found to be the most discriminative both in our study and others [2], [3]. The proposed biometric system was evaluated by the correct recognition rate (CRR) for the identification performance and by equal error rate (ERR) for the authentication. Higher CRR means better performance for identification meanwhile lower ERR implies better performance for authentication.

III. Results and Discussion

A. Identification and authentication performance

Feature extraction routines for HDMRMCL, MCL, BETA, and GAMMA were implemented in Matlab 2018a on a laptop with a 2.7 GHz Intel Core i7 processor. Identification was realized by assigning the signal data to the nearest class, while authentication was carried out with a distance threshold being set to control whether an observation data belongs to that class. Identification and authentication stages were handled using Python 3 with numerical programming routines made publicly available by Yahyei and Özkurt (2022) [9] at https://github.com/RezaYahyaei/Paper2022. The identification and the authentication performances are given in Table I in terms of average classification accuracy with standard deviation for identification and EER for authentication. As

TABLE I: Identification and authentication performance comparison between HDMRMCL, MCL, BETA and GAMMA features

Features	Identification		Authentitcation	
	EO ACC	EC ACC	EO EER	EER
HDMRMCL	100.0 ± 0.0 %	99.7 ± 0.4%	3.47%	7.49%
MCL	99.4 ± 0.4%	98.8 ± 1.1%	6.29%	10.40%
BETA	98.8 ± 0.7%	95.1 ± 1.3%	17.09%	21.71%
GAMMA	95.1 ±1.7%	94.6 ±1.8%	14.92%	19.53%

seen from the table, MCL combined with HDMR provided the highest identification performance compared to only MCL and the other two EEG spectral band based features. Authentication performance is also the highest with EERs of 3.47 % and 7.49 % for the EO and EC conditions, respectively. The superior authentication performances of HDMRMCL can also be observed from Figures 2 and 3.

Table II shows the average runtime measurements of the compared features for 20 runs when data segment length is 10 s. Addendum of HDMR does not increase the computational expense heavily as it takes comparable times with the features derived from spectral band activities. We also investigated the

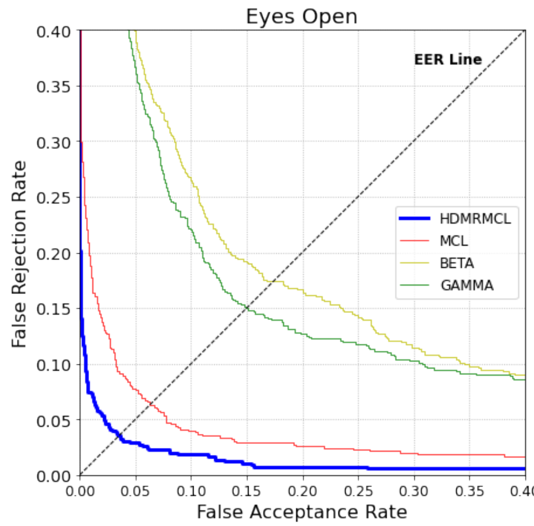

Fig. 2: Detection error trade-off (DET) curves for eyes-open resting state of authentication. The intersection points between the curves and the EER identity line are the equal error rates of the related features.

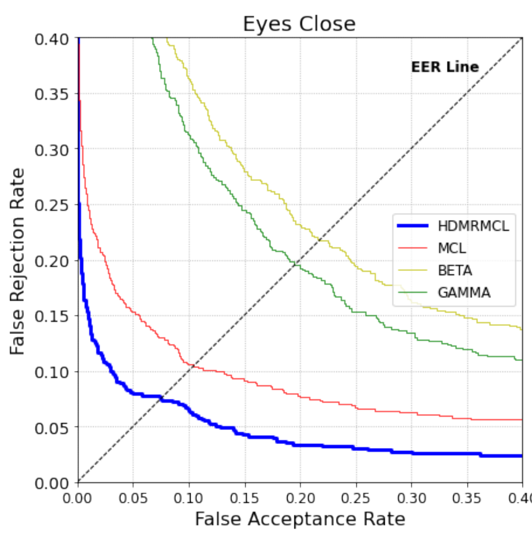

Fig. 3: Detection error trade-off (DET) curves for eyes-close resting state of authentication. The intersection points between the curves and the EER identity line are the equal error rates of the related features.

TABLE II: Running times for the compared features

Feature	Running Time (s)
HDMRMCL	7.4279
MCL	2.3121
BETA	9.3044
GAMMA	7.0820

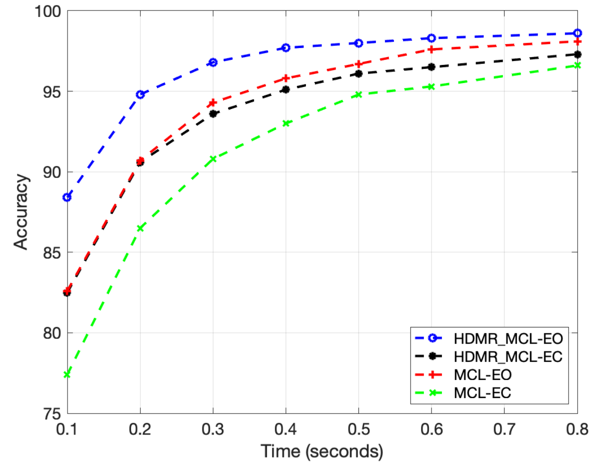

Fig. 4: Identification accuracy change with respect to time less than 1 seconds, $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8\}$ seconds.

identification performance of HDMRMCL for segments of various time lengths. This would yield the temporal sensitivity of the proposed scheme. Figure 4 presents the accuracies with respect to the time segments with duration lengths of less than a second. HDMRMCL provides higher than 90% and 95% accuracies for as short as 0.2 seconds time duration both for EC and EO. For time intervals shorter than 0.2 seconds, an abrupt decrease in performance is apparent. The performances given by HDMRMCL over a second were evaluated for longer time segments with durations of $\{1, 2, 3, 5, 6, 8, 10, 12, 15\}$ seconds. Figure 5 displays the corresponding results compared to only MCL results. HDMRMCL's performance surpasses MCL's significantly for both EO and EC conditions and for all durations.

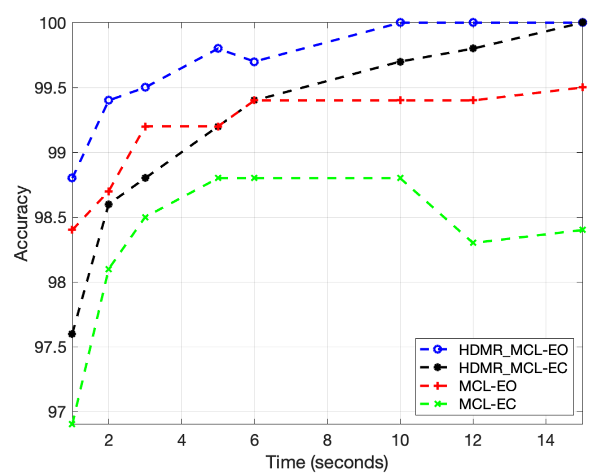

Fig. 5: Identification accuracy change with respect to time more than 1 seconds, $\{1, 2, 3, 5, 6, 8, 10, 12, 15\}$ seconds.

B. Discussion

Results confirmed our hypothesis by showing that the addendum of HDMR to MCL enables distinctive neural fingerprints and offers superior recognition performances when compared to solely MCL and conventional spectral features. Hence, it proved to be a considerable improvement on the biometric system suggested by a recent previous study [9] without incurring significant computational cost. A noteworthy finding of the proposed feature extraction methodology of HDMRMCL is its robustness to different time-lengths as brief as 0.2 s.

Although HDMRMCL has a higher computational complexity with $\mathcal{O}(e \times t \times k)$, ($e$ is the number of channels, t is the number of time points and k is the number of segments) than that of the spectral features of beta and gamma with $\mathcal{O}(e \times t)$, its running time is comparable to them (see Table II). This is because HDMR components are obtained solely through linear multiplications and summations, thereby avoiding the computationally intensive operations within the feature extraction process.

IV. Conclusion

In this study, we proposed a novel feature extraction scheme for EEG-based biometrics named as HDMRMCL. It employs HDMR as an EEG preprocessor and MCL as a feature extractor being fed on a plain Euclidean classifier. HDMRMCL was applied on a standard dataset, that has been numerously exploited in EEG-based biometrics literature. Identification and authentication performances for HDMRMCL were found to be superior compared to those obtained by conventional spectral features and MCL as such. In brief, we demonstrated that the combination of HDMR and MCL offers a rapid, accurate and robust neural biometric recognition method, which satisfies the accuracy and the computational time requirements of a realistic system. A further study would be to test and validate the suggested methodology on larger EEG resting state datasets with a higher number of subjects.

References

[1] G. Niso, E. Romero, J. T. Moreau, A. Araujo, and L. R. Krol, "Wireless EEG: A survey of systems and studies", NeuroImage, vol. 269, pp. 119774, 2023.

[2] J. da Silva Castanheira, H.D. Orozco Perez, B. Misic and S. Baillet, "Brief segments of neurophysiological activity enable individual differentiation", Nature Communications, vol. 12(1), 2021.

[3] K.P. Thomas, A.P. Vinod, "Biometric identification of persons using sample entropy features of EEG during rest state", IEEE Int. Conf. Syst. Man Cybernet, pp. 3487–3492.

[4] Z. Mu, J. Hu, J. Min, "EEG-based person authentication using a fuzzy entropy- related approach with two electrodes", Entropy vol. 18, pp. 432, 2016.

[5] N. A. Alzahab, M. Baldi, and L. Scalise, "Efficient feature selection for electroencephalogram-based authentication", IEEE International Symposium on Medical Measurements and Applications, pp. 1–6, 2021.

[6] L.A. Moctezuma and M. Molinas, "Event-related potential from EEG for a two-step identity authentication system", IEEE Int. Conf. Ind. Inform. vol. 1, 392–399, 2019.

[7] K.Mukaiand and I.Nakanishi, Introduction of fractal dimension feature and reduction of calculation amount in person authentication using evoked EEG by ultrasound, IEEE Region 10 Conference (TENCON), pp. 567–572, 2020.

[8] M. Demuru and M. Fraschini, "EEG fingerprinting: subject-specific signature based on the aperiodic component of power spectrum", Comput. Biol. Med. vol. 120, pp. 103748, 2020.

[9] R. Yahyaei, and T. E. Özkurt, "Mean curve length: An efficient feature for brainwave biometrics", Biomedical Signal Processing and Control, vol. 76, pp. 103664, 2022.

[10] I. M. Sobol, "Theorems and examples on high dimensional model representation", Reliability Engineering and System Safety , vol. 79.2, pp. 187-193, 2003.

[11] Q. Wang and F. Hongbing, "An adaptive high-dimensional model representation method for reliability analysis of geotechnical engineering problems", International Journal for Numerical and Analytical Methods in Geomechanics, vol. 44.12, pp. 1705-1723, 2020.

[12] K. Cheng and L. Zhenzhou, "Time-variant reliability analysis based on high dimensional model representation", Reliability Engineering & System Safety, vol. 188, 310-319, 2019.

[13] G. Taşkın and C. Melba, "Extending out-of-sample manifold learning via meta-modelling techniques", IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017.

[14] X. Yue, J. Zhang, W. Gong, M. Luo & L. Duan, "An adaptive PCE-HDMR metamodeling approach for high-dimensional problems", Structural and Multidisciplinary Optimization, vol. 64, pp. 141-162, 2021.

[15] S. Tuna, E. Korkmaz Özay, B. Tunga, E. Gürvit, and M. A. Tunga, "An efficient feature extraction approach for hyperspectral images using Wavelet High Dimensional Model Representation. International Journal of Remote Sensing", vol. 43(19-24), pp. 6899-6920, 2022.

[16] E. K. Özay, and B. Tunga, "A novel method for multispectral image pansharpening based on high dimensional model representation", Expert Systems with Applications, vol. 170, pp. 114512, 2021.

[17] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals", Circulation, vol 101 (23), pp.e215–e220, 2000.

[18] J.C. Monsy, A.P. Vinod, "EEG-based biometric identification using frequencyweighted power feature", IET Biometrics, vol. 9 (6), pp. 251–258, 2020.

[19] Y. Sun, F. P. W. Lo and B. Lo, "EEG-based user identification system using 1D-convolutional long short-term memory neural networks", Expert Systems with Applications , vol. 125, pp. 259-267, 2019.

[20] M. Ayvaz and M. Demiralp, "Towards a new multiway array decomposition algorithm: elementwise multiway array high dimensional model representation (EMAHDMR)", Proceedings of the 2nd international conference on Applied informatics and computing theory, pp. 76-81, 2011.

[21] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: An efficient feature for seizure onset detection", 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2, pp. 1707–1710, 2001.

[22] T. E. Özkurt, M. Sun, T. Akgül, and R. J. Sclabassi, "Optimal feature selection for seizure detection: A subspace based approach", International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2134–2137, 2006.

[23] L. Guo, D. Rivero, J. Dorado, J.R. Rabunal, A. Pazos, "Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks", Journal of Neuroscience Methods, vol. 191 (1), pp. 101–109, 2010.

[24] L. Logesparan, A.J. Casson, E. Rodriguez-Villegas, "Optimal features for online seizure detection, Medical & Biological Engineering & Computing", vol. 50 (7), 659–669, 2012.