

Linear and Nonlinear Analysis of Intrinsic Mode Function after Facial Stimuli Presentation in Children with Autism Spectrum Disorder

Dalal Bakheet^{a,*}, Koushik Maharatna^a

^a*School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, GB*

ARTICLE INFO

Keywords:

Autism Spectrum Disorder (ASD)
Electro-encephalogram (EEG)
Event-Related Potential (ERP)
Multivariate Empirical Mode Decomposition (MEMD)

ABSTRACT

In this work, a method for classifying Autism Spectrum Disorders (ASD) from typically developing (TD) children is presented using the linear and nonlinear Event-Related Potential (ERP) analysis of the Electro-encephalogram (EEG) signals. The signals were acquired during the presentation of three types of face expression stimuli —happy, fearful and neutral faces. EEGs are first decomposed using the Multivariate Empirical Mode Decomposition (MEMD) method to extract its Intrinsic Mode Functions (IMFs), which provide information about the underlying activities of ERP components. The nonlinear sample entropy (SampEn) features, as well as the standard linear measurements utilizing maximum (Max.), minimum (Min), and standard deviation (Std.), are then extracted from each set of IMFs. These features are then evaluated by the statistical analysis tests and used to construct the input vectors for the Discriminant analysis (DA), Support vector machine (SVM), and k-Nearest Neighbors (kNN) classifiers. Experimental results show that the proposed features can differentiate the ASD and TD children using the happy stimulus dataset with high classification performance for all classifiers that reached 100% accuracy. This result suggests a general deficit in recognizing the positive expression in ASD children. Additionally, we found that the SampEn measurements computed from the alpha and theta bands and the linear features extracted from the delta band can be considered biomarkers for disturbances in Emotional Facial Expression (EFE) processing in ASD children.

1. Introduction

Autism Spectrum Disorders (ASD) is a lifelong condition associated with very high societal costs related to services and lost productivity by patients and their families. As an example, currently, the UK spends around £36B per year only for managing this disorder [1]. ASD is identified by restricted and repetitive behaviour combined with echolalia and different forms of intellectual/motor disabilities. Very early diagnosis (around or before 24–30 months of age) and subsequent early intervention has been shown as the most effective way of treating ASD children in terms of maximal behavioural and intellectual outcomes.


Traditionally, ASD is diagnosed at around 3.5–4 years of age, based on narrative behavioural interactions between the children and specialists besides parental questionnaire. This usually requires an extended period to detect abnormalities, subjective in nature, and lacking from biological evidence besides the fact that such a method cannot be applied for very early diagnosis as for evaluating their behaviour, the children needs to be at a certain age (typically > 3.5 years). Therefore, recently more emphasis has been given to diagnose ASD based on biological markers, thereby increasing evidence-based diagnostic accuracy.

The Electro-encephalogram (EEG) signal is widely used in the literature as an effective tool for diagnostic systems in different brain disorders as it can shed light on the precise temporal dynamics of the brain [2]. Event-Related Po-

tential (ERP) is among the most common patterns extracted from EEG and can help study the signal changes over time in response to an external stimulus. This opens the possibility of diagnosing ASD based on ERP analysis under certain age-appropriate tasks. One such stimulus, where ASD children show severe deficit compared to a typically developing (TD) child, is in Emotional Facial Expression (EFE) processing—a fundamental skill in child development that begins typically in early childhood. Thus, identifying impairments in processing EFE has been suggested as a possible tool for diagnosing ASD children [3]. Monteiro *et al.* [4] have conducted a systematic review of EEG-based ERP studies regarding EFE processing in ASD, and they found that the literature was inconsistent: while some research showed differences in neural responses between ASD and TD individuals, other studies did not identify any deficits. Inconsistency was also found regarding the ASD impairment in recognizing different EFE. Besides, none of the previous explorations considered using the underlying nonlinear dynamics of ERP components in addition to the traditional linear measurements to characterize the signals in response to EFE stimuli in individuals with ASD.

Nonlinear features consist of quantitative measures representing the complex dynamic characteristics of the EEG signals, which the linear measurements cannot capture. In this work, we propose including the nonlinear features of ERP components with the traditional linear ones to classify ASD and TD children. More precisely, we employ the sample entropy (SampEn) [5] as a nonlinear measure to detect the signal's irregularity in response to the EFE stimuli. Simultaneously, three standard linear features, namely maximum (Max.), minimum (Min), and standard deviation (Std.), were used to detect the amplitude changes after EFE presen-

*Corresponding author

 dmb1y17@soton.ac.uk (D. Bakheet)

ORCID(s): 0000-0003-2227-6801 (D. Bakheet); 0000-0001-6518-2755

(K. Maharatna)

¹Computer Science and Artificial Intelligence, University of Jeddah, 21959 Jeddah, SA

tation.

Before computing the measurements above, the Multi-variate Empirical Mode Decomposition (MEMD) method [6] has been used to decompose EEG signals into a set of scales, namely Intrinsic Mode Functions (IMFs), and subsequently, extracting the features from each of these IMFs. MEMD was employed here as the SampEn may fail to account for time series with multiple time scales [5]. Besides, extracting the features from the signal's intrinsic components gives insight into the overlapping time-frequency activity underlying ERP components. Contrary to other decomposition approaches, the MEMD method adaptively decomposes the signals and does not require a priori selection of the filter cut-offs. This characteristic solves the well-known variability between subjects in the frequency ranges of the traditional brain waves, leading to missing potentially meaningful brain dynamics.

It is the first time such an approach involving both linear and nonlinear intrinsic characteristics of the ERPs has been employed for ASD diagnosis to the best of our knowledge. Statistical analysis is initially utilized to evaluate the proposed features. The features were then used to train and test three well-known classifiers, namely Discriminant analysis (DA), Support vector machine (SVM), and k-Nearest Neighbors (kNN), to show how useful they could be in practical ASD classification.

Our quest here could be summarized for exploring the answers to a set of fundamental questions: (1) Is the proposed combination of linear and nonlinear features can classify ASD and TD? (2) Which IMF component is the best to reveal ASD brain abnormality, and what is its underlying frequency? (3) Which facial expression stimulus-evoked ERP response gives better discrimination between ASD and TD children?

The remainder of this paper is organized as follows: Section 2 describes the EEG dataset used in this work, followed by brief descriptions of the SampEn, MEMD method and the feature extraction scheme adopted here to extract the linear and nonlinear measures from the IMF components. Descriptions of the statistical analysis, feature selection and the classification procedure are also provided in the same section. The results are presented in Section 3 and discussed in detail in Section 4. Section 5 concludes the paper and suggests some of the future research directions.

2. Materials and methods

In this study, we used an available EEG dataset recorded during the presentation of EFE to explore whether the proposed features in the MEMD domain can efficiently discriminate between ASD and TD children. Thus, after decomposing the signals using the MEMD method, the proposed linear and nonlinear features are computed from selected IMFs and used to form feature vectors. The entire set of features and the selected ones (using a feature selection technique) were used to train and test DA, SVM, and kNN classifiers for ASD classification. Figure 1 presents the overall process

of the proposed framework. The whole analysis is carried out in the MATLAB software package R2018a.

2.1. Experimental data description

The dataset was taken from a previous EEG-based EFE ERP studies [7, 8]. The dataset contains EEG signals from 24 subjects —12 ASD and 12 TD; age group 6–13 years (mean age 10.2 and 9.7 years for ASD and TD, respectively).

The EEG signals were acquired during the presentation of three kinds of EFE —neutral, happy and fearful. The experiment was done in 4 blocks, and in each block, 10 neutral, 10 happy and 10 fearful faces were presented twice at random order. Signals were recorded at 250 Hz using a 128-channel HydroCel Geodesic Sensor net, as illustrated in Figure 2. The acquired EEG data were segmented into 1000 ms epochs (150 ms baseline and 850 ms post-stimulus presentation) to focus on the time window surrounding the actual event. Epochs with signals over a threshold of 200 μV were considered artefacts and rejected. Data were band-pass filtered with cut-off frequencies from 0.5 Hz to 50 Hz to remove low-frequency drifts and high-frequency measurement noise using a fifth-order forward-backward Butterworth filter, and the baseline was corrected [7, 8].

Eight channels located in the Fusiform Gyrus (FG) area of the cortex are selected for the current study. FG is believed to be the specific region for processing face features and emotions [7]. Right Region Of Interest (ROI) was defined as electrodes no. 96 (P8/T6), 95 (P10), 90 (PO8) and 89, whereas left ROI as electrodes no. 58 (T5), 64 (P9), 65 (PO7) and 69. These channels are highlighted in red in the electrode arrangement in Figure 2.

2.2. Sample entropy (SampEn)

SampEn, developed by Richman and Moorman [5], effectively measures the complexity of nonlinear physiological signals such as EEG. SampEn is a modification of approximate entropy [9], improving its computation and accuracy of signal regularity. SampEn has defined as the probability that two similar patterns for m point remain identical at the next $m+1$ point within a tolerance r . For the time series $x(i)$ of length N , SampEn is given by:

$$SampEn(m, r, N) = -\ln[A^m(r)/B^m(r)], \quad (1)$$

where

$$A^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} C_i^{m+1}(r), \quad (2)$$

$$B^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} C_i^m(r), \quad (3)$$

$$C_i^m(r) = (N - m - 1)^{-1} C_i, i = 1, 2, \dots, N - m, \quad (4)$$

where m is the embedding dimension, $B^m(r)$ is the likelihood that $X_m(i)$ and $X_m(j)$ is matching for m points, while $A^m(r)$ is the likelihood that $X_m(i)$ and $X_m(j)$ will match for

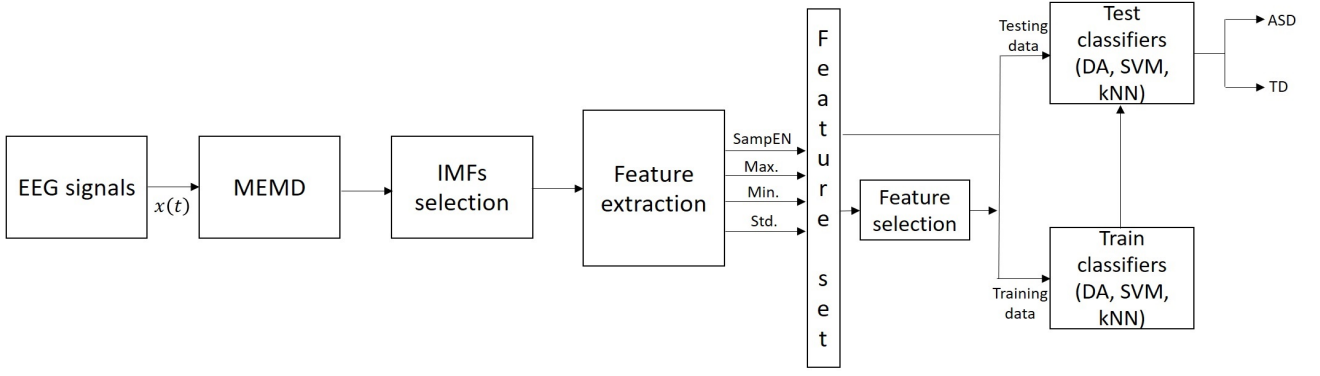


Figure 1: Block diagram of the proposed method.

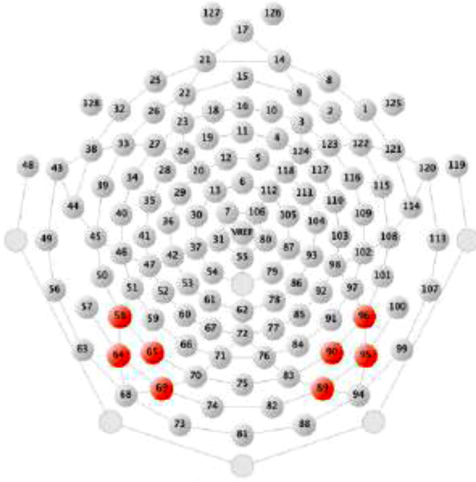


Figure 2: Right and Left ROIs used in the present study [7].

$m+1$ points. $C_i^m(r)$ is the probability of a vector $X_m(i)$ being similar to $X_m(j)$ within a tolerance r , C_i is the number that the distance of two vectors $X(i)$ and $X(j)$ is smaller than r , and a vector $X_m(i) (1 \leq i \leq N - m + 1)$ is reconstituted of this series, and is given by: $X_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\}$. To estimation SampEn optimally, some studies have recommended the embedding dimension $m = 2$ or 3 , and the tolerance $r = 0.1 - 0.25$ of the standard deviation of the signal [10, 11]. In this exploration, we set $m = 3$, and $r = 0.2$ of the standard deviation for the processed signal.

Given the multiple tempo-spectral scales inherent in the brain, SampEn analysis is not appropriate to be applied directly to the EEG signals [5]. Thus, EEG needs to be decomposed first into single scale components to measure its underlying nonlinear dynamic. The MEMD method has been employed for this purpose.

2.3. Multivariate empirical mode decomposition (MEMD)

MEMD was introduced by Rehman *et al.* [6] as a multivariate extension of the Empirical Mode Decomposition (EMD) [12] method to resolve the mode alignment prob-

lem. EMD and their extensions have recently become useful tools for the time-frequency analysis. Unlike other time-frequency methods such as Short-time Fourier (STFT) [13] and Wavelet transforms [14], EMD-based methods do not require predefined frequency bands to analyse EEG signals. Thus, they decompose the time-series adaptively, through the Sifting process, from high to low-frequency components known as IMFs. The resulted IMFs should sum to a composite nearly identical to the original signal given by:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t). \quad (5)$$

where, x is the original signal, i being indices of the IMFs, n is the total number of IMFs, r is the residue at the end of the Sifting process [12]. The procedure of the Sifting process of the MEMD method starts by considering a sequence of n -dimensional vectors:

$\{v(t)\}_{t=1}^T = \{v_1(t), v_2(t), v_3(t), \dots, v_n(t)\}$ that represents a multivariate signal with n components, and $X^{Q_k} = \{x_1^k, x_2^k, x_3^k, \dots, x_n^k\}$ denoting a set of direction vectors along the directions given by angles $Q^k = \{Q_1^k, Q_2^k, \dots, Q_{(n-1)}^k\}$ on an $(n-1)$ -sphere. Then the MEMD algorithm is summarized as following:

- 1) Choose a suitable set of points for sampling on a $(n-1)$ sphere.
- 2) Calculate a projection, denoted by $\{P^{Q_k}(t)\}_{t=1}^T$, of the input signal $\{v(t)\}_{t=1}^T$ along the direction vector X^{Q_k} , for all k (the whole set of direction vectors), giving $\{P^{Q_k}(t)\}_{k=1}^K$ as the set of projections.
- 3) Find the time instants $t_j^{Q_k}$ corresponding to the maxima of the set of projected signals $\{P^{Q_k}(t)\}_{k=1}^K$.
- 4) Interpolate $[t_j^{Q_k}, v(t_j^{Q_k})]$ to get the multivariate envelope curves $\{e^{Q_k}(t)\}_{k=1}^K$.
- 5) For a set of K direction vectors, the mean $m(t)$ of the envelope curves is calculated as $m(t) = \frac{1}{K} \sum_{k=1}^K e^{Q_k}(t)$.
- 6) Extract the detail $c_i(t)$ using $c_i(t) = v(t) - m(t)$ (i is an order of IMF). If the detail $c_i(t)$ satisfies the IMF conditions, apply the above procedure to $v(t) - c_i(t)$, otherwise apply it to $c_i(t)$.

The Sifting process can be stopped when the detail $c_i(t)$ is monotonic and no more IMFs can be extracted from it.

Once all IMFs are identified, the instantaneous frequencies of each IMF can be acquired by the Hilbert Transform (HT) [12]. MEMD also works as a filter where it can isolate the inevitable noise inherent in the time series in separate components. Thus, after decomposing the signal, the noisy components can be identified and removed [6].

2.4. MEMD-based feature extraction process

To characterize the multi-subject neural recordings collected from multiple channels over multiple trials, our proposed MEMD-based analysis consists of the following steps:

- 1) The epochs of each class of stimulus are averaged at the beginning of this analysis to enhance the signal to noise ratio (SNR) and extract the ERP components.
- 2) For each class of stimulus, the data points of all subjects from each channel are stacked on top of each other. Hence, we construct eight matrices (one for each channel); each of them has the dimensionality of $N_s \times N_t$, where N_s denotes the number of subjects (which is 24) and N_t indicates the number of temporal samples (which is 250). Figure 3 illustrates this process.
- 3) The MEMD method is then applied to each matrix separately, as shown in Figure 3. Following this way, the datasets of all subjects for a specific channel are decomposed into the same number of IMFs.
- 4) As the number of IMFs may vary among channels, the lowest number of modes can be considered. In the data under study, the channels decomposed into the same number of IMFs, which was eight for each class of stimulus. Figure 4 gives an example of the resulted IMFs from channel 1 of the first ASD subject collected during the happy stimulus presentation.
- 5) The frequencies of each IMF is then acquired by HT. It was found that IMF1 and IMF2 are noisy and contain different oscillatory components. Therefore, these modes were excluded from further analysis. IMF8 was also ignored as it represents the residue mode, which might give unreal information about the signal. The frequencies of the remaining IMFs were localized approximately in the following ranges: IMF3 (30–37 Hz), IMF4 (13–20 Hz), IMF5 (8–12 Hz), IMF6 (4–6 Hz), and IMF7 (0.5–2.5 Hz). According to the traditional ranges of the five physiological frequency bands [15], it was found that IMF3 to IMF7 frequencies belong to the gamma, beta, alpha, theta and delta band, respectively.
- 6) SampEn are then computed over each IMF (IMF3–IMF7) to depict their complexity in the time domain.
- 7) Standard linear features (Max., Min. and Std.) are also extracted from each IMF (IMF3–IMF7) to reflect the temporal changes of the amplitude after the stimuli presentation.

2.5. Statistical analysis

Statistical analysis test was used to determine the capability of the proposed set of features to discriminate between

the two classes (ASD and TD) and to choose the statistically significant ones for classification purposes. To this end, we applied the one-way analysis of variance (ANOVA) using MATLAB's statistics toolbox.

Generally, statistical tests include computing test statistics translated as a statistically significant or non-significant value greater or less than a threshold known as the level of significance (α). Being without apparent reason, the most common value of α is 0.05 [16]. This value of α needs to be adjusted (lowered) mainly when several independent tests are being performed simultaneously as the chosen value of α may be appropriate for each test, but not for the set of all tests. Bonferroni correction [17] is the straightforward approach for such adjustment, in which the given α is divided by the total number of running tests to find a corresponding level of significance. In this study, the discriminative capability of each feature (SampEn, Max., Min., and Std.) computed from each IMF has been evaluated separately, and this process was performed simultaneously for each type of stimuli. Hence, Bonferroni correction has been employed, and the value of α has been corrected from 0.05 to 0.0008. Thus, a difference is statistically significant if $p\text{-value} < \alpha (= 0.0008)$.

2.6. Features selection

The proposed method uses four features extracted from the multivariate IMF signal of each EEG channel. Thus, the feature's number could be large and lead to the overfitting problem. To prevent overfitting, the number of features should be relatively small with respect to the number of training samples to ensure good generalization performance of the designed classifier. Thus, selecting highly informative features from a larger pool of available ones is crucial to optimize the classification performance [18].

Two different feature selection techniques are known in the machine learning literature—Scalar feature selection and feature vector selection [18]. Scalar feature selection is employed independently of the classifier. The features are ranked in descending order using a score like Fisher's Discriminant Ratio (FDR). The top-ranked features are then selected to check a particular classifier's performance [18].

On the other hand, feature vector selection identifies the best combination of features based on several search techniques, such as sequential backward search (SBS) and sequential forward search (SFS) [18]. The advantage of the scalar feature selection over the feature vector selection is computational simplicity while achieving the ultimate goal of getting a reliable classification [18]. Thus, scalar feature selection method based on FDR criterion is employed in this study to select the highly discriminant features as follows:

- 1) Normalize the features to zero mean and unit variance to remove the bias from features having high values.

$$\hat{x}_i = \frac{x_i - \bar{x}}{\sigma}, i = 1, 2, \dots, N \quad (6)$$

where \hat{x}_i is normalized value, N be the number of features, x_i is the feature i , \bar{x} the mean and σ be the standard deviation.

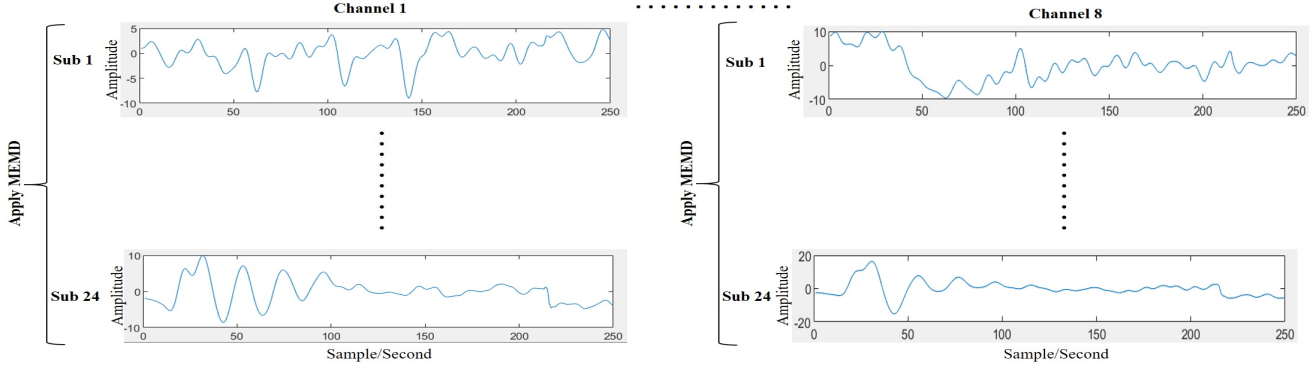


Figure 3: Simultaneous decomposition of the dataset recorded during the happy stimulus presentation.

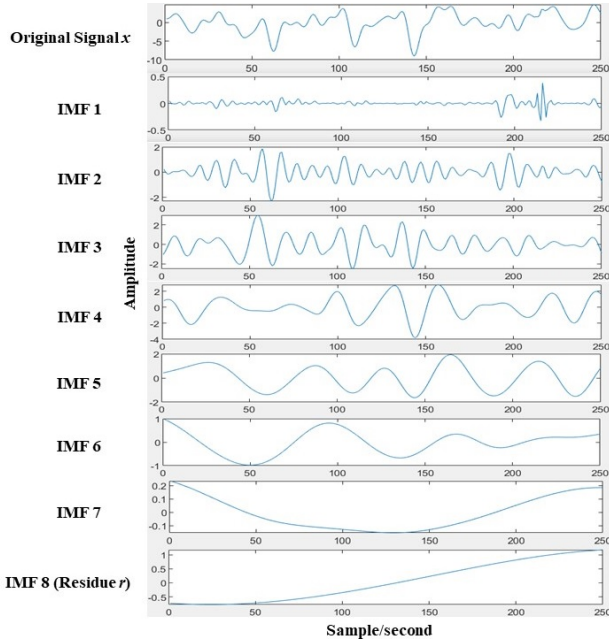


Figure 4: An example of the resulted IMFs from the MEMD method.

- Rank the features in descending order according to the FDR measure.

$$FDR = \frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)} \quad (7)$$

where μ_1 is the mean of the first class, μ_2 is the mean of the second class, σ_1^2 and σ_2^2 are the variance of the first and second class, respectively.

- Compute the cross-correlations among the top-ranked feature (with the index i_1) and each remaining features. The index, i_2 , of the second most important feature is computed as

$$i_2 = \operatorname{argmax}\{a_1 C_j - a_2 |P_{i_1, j}|\}, j \neq i_1 \quad (8)$$

which incorporates the feature ranking value C for the j th feature, and the cross-correlation ($P_{i_1, j}$) between the

best feature (i_1) and feature $j \neq i_1$. The parameters a_1 , a_2 are weighting factors.

- The rest of the features are ranked according to

$$i_k = \operatorname{argmax}\left\{a_1 C_j - \frac{a_2}{k-1} \sum_{r=1}^{k-1} |P_{i_r, j}|\right\}, j \neq i_r \quad (9)$$

for $r = 1, 2, \dots, k-1$, and $k = 3, 4, \dots, m$. The average correlation with all the previously considered features is taken into account [18].

2.7. Machine learning

To assess our proposed features for the practical ASD classification, Classification Learner App within the Statistics and Machine Learning Toolbox in MATLAB was used to train DA, SVM, and kNN classifiers. For DA, two popular discriminant functions were investigated—linear (LDA) and quadratic (QDA). Three kernel functions were utilized for SVM—linear (L-SVM), quadratic (Q-SVM) and cubic (C-SVM). For kNN, a different number of nearest neighbor k was used, and the best accuracy was reported. For a detailed review of these classifiers, [19] can be consulted.

In order to obtain the classifiers' performance, we randomly selected 75% of the total number of samples to be used for training, and the remaining 25% (which were unseen by the models during the training) were used for testing the resulted model. The 10-fold cross-validation technique was used in the training process to prevent overfitting due to its reliability and lowest chance of introducing an undesired bias with few samples [20]. Thus, 90% of the training data samples were used for training, and 10% of the hold-out samples were used for validation. This procedure was repeated for ten folds, and the classifier performance was then obtained by averaging the ten independent results. During the validation process, the hyperparameters were tuned using hyperparameter optimization within the Classification Learner app. The app tries different combinations of hyperparameter for each model type using an optimization scheme that seeks to minimize the model classification error. Table 1 shows the division of the dataset for the training, validation, and testing process.

Table 1

Details of datasets division into training and testing sets.

Division	Training/validation	Testing
Percentage	75% (90% for training and 10% for validation)	25%
Samples	18 (16 for training and 2 for validation)	6

Table 2

p-values of the proposed features for the happy, fear, and neutral stimuli datasets. Features that are statistically significant are indicated in boldface.

Component	Feature	Happy	Fear	Neutral
IMF3	SampEn	0.06	0.43	0.82
	Max.	0.81	0.84	0.43
	Min.	0.44	0.54	0.78
	Std.	0.73	0.85	0.57
IMF4	SampEn	0.08	0.81	0.58
	Max.	0.01	0.32	0.37
	Min.	0.06	0.47	0.79
	Std.	0.01	0.36	0.34
IMF5	SampEn	0.000007	0.04	0.00001
	Max.	0.73	0.57	0.66
	Min.	0.57	0.29	0.53
	Std.	0.45	0.49	0.91
IMF6	SampEn	0.00002	0.16	0.84
	Max.	0.25	0.70	0.77
	Min.	0.18	0.14	0.31
	Std.	0.62	0.54	0.76
IMF7	SampEn	0.05	0.80	0.48
	Max.	0.01	0.05	0.004
	Min.	0.0007	0.02	0.001
	Std.	0.0005	0.02	0.002

The classification performance was evaluated using the conventional measures of accuracy (ACC), sensitivity or true positive rate (TPR), and specificity or true negative rate (TNR).

3. Results

3.1. ANOVA results

Table 2 presents the p-values of the proposed features for the happy, fear, and neutral stimuli datasets. It has been noticed that SampEn features computed from IMF5 significantly differ between the ASD and TD groups in the happy and neutral datasets. SampEn extracted from IMF6 in the happy stimulus case also gave a low p-value indicating good discriminative capability between the two groups. The results also indicate that the Min. and Std. features extracted from IMF7 can significantly differentiate between ASD and TD signals in happy stimulus case. All features computed from IMF3 and IMF4 gave high p-values indicating low discriminatory capability. Thus, IMF3 and IMF4 were excluded from the classification process. Besides, Max. feature also achieved a high p-value in all comparisons. Therefore, it was eliminated from the training and testing matrices before feeding them into the classifiers.

Table 3

Mean values of the channel-based SampEn features computed from IMF5 (for the happy and neutral stimuli) and IMF6 (for the happy stimulus) in each group (ASD and TD).

Channel no.	Group	IMF5		IMF6
		Happy	Neutral	Happy
96 (P8/T6)	TD	0.231891	0.255047	0.179828
	ASD	0.180592	0.195368	0.162561
95 (P10)	TD	0.237618	0.251609	0.165598
	ASD	0.177465	0.191654	0.157265
90 (P08)	TD	0.252373	0.240605	0.192727
	ASD	0.216844	0.206553	0.187306
89	TD	0.238013	0.242491	0.170675
	ASD	0.192634	0.183182	0.158942
58 (T5)	TD	0.253348	0.212844	0.18533
	ASD	0.203716	0.2226	0.157741
64 (P9)	TD	0.211596	0.213144	0.17435
	ASD	0.187116	0.199813	0.161868
65 (P07)	TD	0.257222	0.238206	0.224777
	ASD	0.257747	0.220763	0.177993
69	TD	0.233715	0.243575	0.193015
	ASD	0.182412	0.204113	0.159689

3.2. SampEn analysis

To show how the complexity differs between ASD and TD children, we computed the mean values of SampEn features in each group (ASD and TD) that are extracted from IMF5 (for the happy and neutral stimuli) and IMF6 (for the happy stimulus) for all channels. Table 3 presents the results. Compared to the TD group, the ASD group showed lower SampEn in almost all FG channels during EFE processing tasks, indicating a complexity reduction of the FG brain activity, specifically in the alpha and theta bands—the frequency bands corresponding to IMF5 and IMF6, respectively.

3.3. Machine learning results

DA, SVM, and kNN classifiers were trained on different dimensions of feature vectors to find the optimal pool of features that can best distinguish between the ASD and TD children. The dimension of the vectors was as follows: Each feature from IMF5, IMF6 and IMF7 was first evaluated for classification separately. Thus, the feature vectors for each subject were $N_e \times N_f \times N_m$, where N_e is the number of channels, N_f is the number of features, and N_m is the number of IMFs, i.e. ($8 \times 1 \times 1 = 8$). The classifiers were then trained with the combination of the three features (SampEn, Min. and Std.) that computed from each IMF, and the dimensions of the feature vectors in this case were ($8 \times 3 \times 1 = 24$).

Table 4
L-SVM performance using different features vectors for all types of stimuli.

Component/s	Feature/s	Happy			Neutral			Fear		
		ACC	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR
IMF 5	SampEn	72.2%	78%	67%	72.2%	78%	67%	61.1%	78%	44%
	Min.	55.6%	67%	44%	38.9%	56%	22%	55.6%	89%	22%
	Std.	50%	67%	33%	44.4%	67%	22%	55.6%	78%	33%
	All	77.8%	89%	67%	72.2%	78%	67%	61.1%	78%	44%
IMF 6	SampEn	77.8%	78%	78%	66.7%	67%	67%	50.0%	67%	33%
	Min.	61.1%	44%	78%	72.2%	67%	78%	61.1%	44%	78%
	Std.	55.6%	44%	67%	55.6%	44%	67%	61.1%	89%	33%
	All	88.9%	89%	89%	72.2%	67%	78%	38.9%	33%	44%
IMF 7	SampEn	55.6%	78%	33%	44.4%	67%	22%	55.6%	44%	67%
	Min.	77.8%	67%	89%	66.7%	67%	67%	61.1%	44%	78%
	Std.	77.8%	56%	100%	72.2%	67%	78%	61.1%	44%	78%
	All	72.2%	67%	78%	66.7%	56%	78%	61.1%	44%	78%
IMF 5-7	SampEn	88.9%	100%	78%	66.7%	78%	56%	44.4%	56%	33%
	Min.	77.8%	67%	56%	61.1%	56%	67%	61.1%	56%	67%
	Std.	83.3%	67%	100%	66.7%	56%	78%	55.6%	56%	56%
	All	94.4%	89%	100%	72.2%	67%	78%	61.1%	44%	78%

Additional investigations were obtained to explore whether combining IMF5, IMF6, and IMF7 could improve the classifiers' performance. In this case, the dimensions of the feature vectors were ($8 \times 1 \times 3 = 24$) to explore each feature individually and ($8 \times 3 \times 3 = 72$) for the combination of the three features (i.e. SampEn, Min. and Std.).

In general, the 10-fold cross-validation results of all classifiers showed an enhancement in the classification performance when the classifiers were trained on the combination of the three features extracted from the last three IMFs (i.e. on the feature vector of size 72). Table 4 gives the detailed validation results of L-SVM to show an example of how the performance improved by combining all features, especially with the happy dataset. All other classifiers had similar improvement.

The classifiers were then trained on different subsets of the 72 features to prevent overfitting. The features on each subset were selected based on the scalar feature selection procedure described previously. In practice, one has to experiment with a different number of selected features and choose the one that results in the best classification performance [18]. Thus, we started by exploring the discrimination capability of the top high ranked 15 features, where the role of thumb is not to exceed the number of the training samples (which is 18) [18]. The top 10 and top 5 features were also investigated. Figure 5 gives the 10-fold cross-validation accuracies for all three stimuli when the classifiers were trained on the entire feature set and the top selected ones. Generally, it can infer that the classifiers' performance was enhanced by reducing the number of features in the happy, neutral, and fear datasets. The best classification accuracies were when the top 10 selected features are used to train almost all the classifiers. Visualizations of the training sets associated with the full and the reduced set of features for all stimuli are also shown in Figure 6. As seen,

the two groups can often better separated, especially in the happy dataset, using the reduced set of the highly selected features.

It has also been noticed from Figure 5 that the best discrimination between ASD and TD children, for all classifiers, were achieved using the happy dataset. The same figure also shows that the classifiers' results for each stimulus type were close to each other, and the QDA achieved the worst performance in general. The best accuracy for the happy dataset was 100% when the top 10 or 15 features were used to train the LDA, kNN and all SVMs. The best accuracy was 88.9% for the neutral dataset when the top 10 or 15 features were used to train the LDA, kNN, L-SVM and Q-SVM. The best accuracy the fear dataset was 83.3% when the top 10 features were used to train the kNN and all SVMs.

The trained models with the top 10 selected features were then used to make predictions with the test datasets. This set of features were employed due to its best and consistent validation performance of all classifiers with the three stimuli datasets. Table 5 gives the testing results. Again the happy dataset had the best classifiers' performance that reached 100% ACC, 100% TPR, and 100% TNR using the LDA, and Q-SVM. Followed by the fear dataset with 50% ACC, 100% TPR, and 0% TNR using Q-SVM and 50% ACC, 66.7% TPR, and 33.3% TNR using QDA. The best performance for the neutral dataset was 33.3% ACC, 0% TPR, and 66.7% TNR using LDA and kNN. The confusion matrix of the best testing performance for each dataset is shown in Figure 7.

4. Discussion

This study addresses a multimodal abnormality in EFE processing in ASD children over the FG brain region. It investigates the impairments in the intrinsic EEG complexity (as indexed by SampEn measures) and the ERP alter-

Linear and Nonlinear Analysis of Intrinsic Mode Function

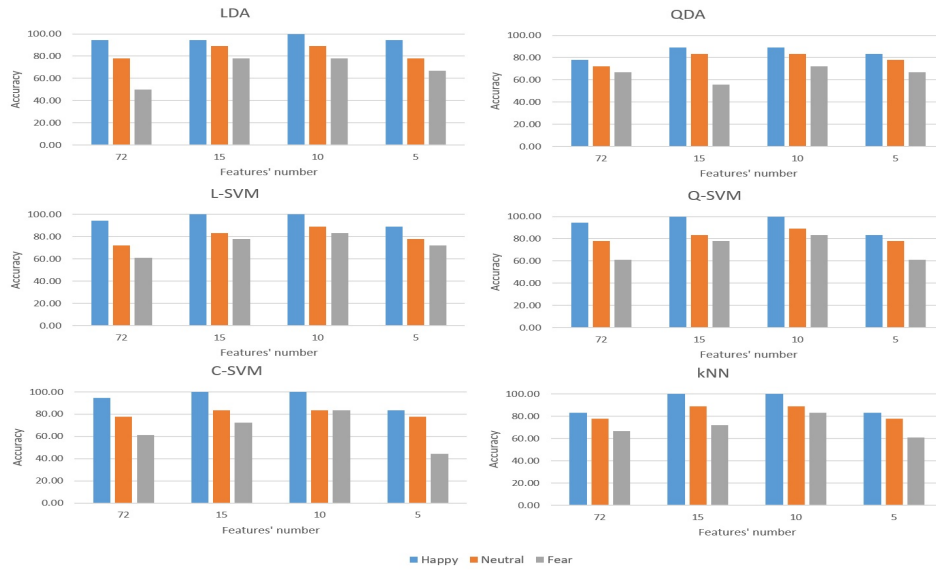


Figure 5: Accuracy of different classifiers with different group of features for each stimulus.

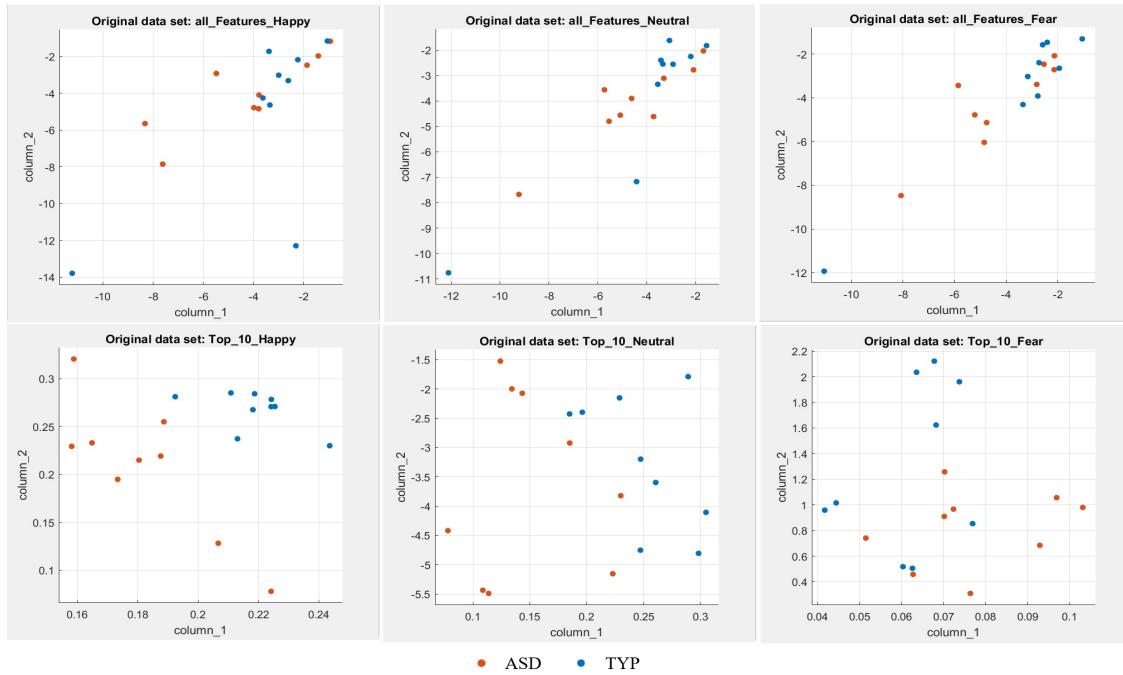


Figure 6: Visualizations of the training sets associated with the full and the top 10 sets of features for all stimuli.

Table 5

Prediction performance of all classifiers on the testing data for all stimuli.

Classifier	Happy			Neutral			Fear		
	ACC	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR
LDA	100%	100%	100%	33.3%	0%	66.7%	33.3%	33.3%	33.3%
QDA	66.7%	100%	33.3%	16.7%	0%	33.3%	50.0%	66.7%	33.3%
L-SVM	83.3%	66.7%	100%	16.7%	0%	33.3%	33.3%	33.3%	33.3%
Q-SVM	100%	100%	100%	16.7%	0%	33.3%	50.0%	100%	0%
C-SVM	83.3%	100%	66.7%	16.7%	0%	33.3%	33.3%	66.7%	0%
kNN	83.3%	100%	66.7%	33.3%	0%	66.7%	33.3%	33.3%	33.3%

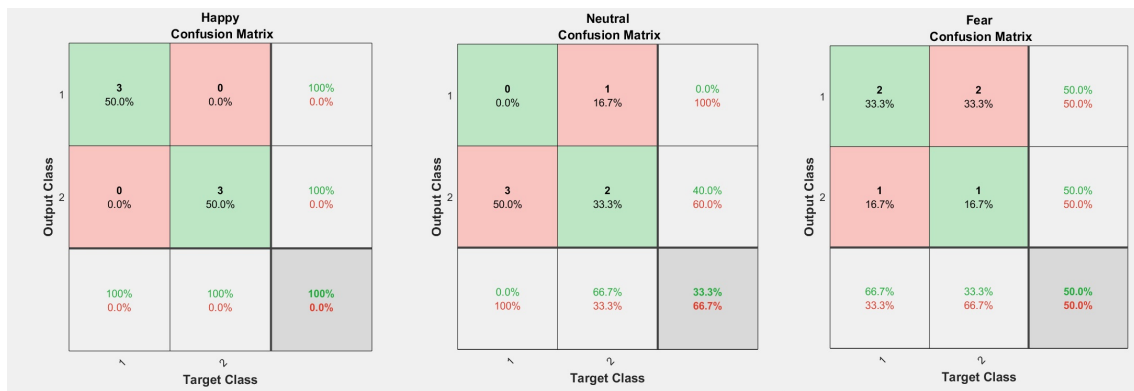


Figure 7: Confusion matrix of the best classification results for each stimulus. Here class 1 depicting ASD, and class 2 representing TD.

ations (through the standard linear features). The statistical analysis tests showed significant differences between ASD and TD children, especially in processing the happy stimulus through the SampEn features computed from IMF5 and IMF6. IMF5 and IMF6 were found (through the HT analysis) to contain the alpha and theta bands, respectively.

Particularly, our results showed a decrease in EEG complexity in the ASD group compared to the TD group. This result is in line with Catarino *et al.* [21] study, which demonstrated a reduction of EEG signal complexity in the ASD group in response to a visual matching task. Hence, our findings support the hypothesis that the complexity of electrical brain activity is reduced in individuals with ASD due to EEG long-range temporal correlations reduction and atypical neural connectivity in the ASD population [21]. The impairments in the alpha band measurements of ASD signals in response to EFE stimuli have also been reported previously in the literature. Khan *et al.* [22] measured functional brain connectivity of magnetoencephalography (MEG) signals in the FG region while subjects were doing emotional recognition tasks. They found that children with ASD demonstrated specific alpha abnormalities compared to TD ones.

Statistical analysis tests also showed significant differences between the two groups in processing the happy stimulus through the linear measurements computed from IMF7, which was found (through the HT analysis) to include the delta band's frequencies. This finding indicates that the features extracted from the low-frequency component could significantly account for the linear ERP variations between the two groups. The role of the delta band in recognition of EFE has been reported before in the literature [23, 24]. In particular, Dominguez *et al.* [25] found that those with ASD had increased coherence in the lowest frequency bands, delta and theta, over occipital channels, including the FG region, as viewing emotional faces compared to controls.

We also trained and tested different classifiers with various feature vectors constructed from IMF5-IMF7 to investigate their practical capability in ASD diagnosis. Almost all validation results suggested that combining the nonlinear measurements (as indexed by SampEn) and the linear features (Min. and Std.) and extracting them from the combina-

tion of the last three IMFs (to include alpha, theta, delta brain bands) can improve the classifiers' performance. Besides, selecting the highly informative features from this large pool using the scalar feature selection technique was further enhanced the results.

In general, the results of all employed classifiers were comparable in each case of stimuli during both the validation and testing procedures. The results showed that the signals recorded during the happy stimulus presentation achieved the best classification performances. This finding suggests general impairment in the positive emotions recognition (as indexed by happiness) in children with ASD. This result is consistent with Yeung *et al.* [26] finding, but it contradicts other studies that only found a deficit in recognition of negative emotions (such as fearful faces) [27, 28]. This inconsistency might be due to the differences in the task demands that widely vary across studies [26]. Nevertheless, the testing results of the fear dataset indicate better performance than the neutral stimulus results. More precisely, the TPR was better in all classifiers. A good TPR is often important in a diagnostic test where children identified as having ASD should be highly likely to have the condition.

The results of this study are first compared to three previously published works that used the same dataset employed herein. Apicella *et al.* [7] assessed the mean values of the ERP latencies and amplitudes using the ANOVA repeated measures of the three stimuli. They reported reduced amplitudes and delayed latencies of all the early ERPs in ASD children regardless of the EFE stimuli. Thus, they suggested that ASD children similarly process all the three faces compared to TD ones. However, they neither studied the frequency alteration nor nonlinear activity underlying ERP components. Jamal *et al.* [8] reached 94.7% ACC as the best performance with 85.7% TPR and 100% TNR using functional connectivity features and the SVM. Although these promising results, their exploration mainly combined all data from the three stimuli. Thus, they tested the capability of the connectivity features to differentiate between the ASD and TD children without giving attention to the effect of each EFE (happy vs. fear vs. neutral). Though their features were based on time-frequency analysis, they also did not discuss the oscil-

lations of interest associated with their findings. Khuntia *et al.* [29] carried out multivariate pattern analysis using Class-wise Principle Component Analysis in both time and time-frequency domains (via STFT). Classification performance reached around 81% in the time domain analysis and 84% in the time-frequency domain exploration regardless of the face and non-face stimuli (as they include the tree stimulus). In the time-frequency analysis, alpha and beta oscillations seemed to identify ASD children best. Nevertheless, this work also did not consider the nonlinear behaviour of the EEG signals.

We also compared our study with two relevant EEG-based ERP studies that employed another children dataset for ASD diagnosis. Yeung *et al.* [26] employed happy, sad, anger, disgust, fear, surprise, and neutral stimuli, and their samples included 18 TD and 18 ASD children ranging in age from 9–10 years-old. They used ANOVA repeated measures to assess functional connectivity, as indexed by theta coherence, during EFE recognition tasks. They found that ASD children exhibited abnormal patterns in the theta coherence associated with their EFE recognition ability. Their results also suggested general impairment in recognition of the positive emotions. Despite their significant findings, they only studied theta wave's connectivity and did not consider the nonlinear activity underlying ERP components. Dominguez *et al.* [25] used fearful and happy stimuli, and their dataset consists of 31 TD and 72 ASD children (age ranges 2 to 4 years 11 months). They trained and tested L-SVM with high dimensional feature vectors of the imaginary part of coherency of all frequency bands, and classification accuracy of 80% was reached. They also carried out ANOVA analysis and found that ASD children exhibited enhanced synchronization during the post-stimulus time, particularly at lower frequency bands. However, they did not report the differences in processing the happy and fearful expressions and did not consider the nonlinearity of the signals. Table 6 summarizes all this researches.

From the above studies and the systematic review of Monteiro *et al.* [4], which review other 14 state-of-the-art articles, we can argue that our exploration's novelty is mainly represented by including linear and nonlinear characteristics of ERPs. Almost all of the reviewed studies did not consider the nonlinearity of the EEG signals when analysing the ERP components. Additionally, even though few articles studied the overlapping time-frequency activity in response to the recognition of emotional facial expression tasks [25, 26, 29], they often spectrally analysed the signals using time-frequency methods that rely on the predefined traditional brain waves. The prior selection of the filter cut-offs raises an issue due to the well-known variability between subjects in the neural oscillations associated with different tasks. This limitation has been settled in our proposed approach using the MEMD method, which decomposes the signals adaptively. Hence, all potentially meaningful brain dynamics are included in the analysis.

Overall, the proposed framework successfully diagnosed ASD with high efficiency that outperforms the results of the

related works. Therefore, it could be instrumental in solving the late diagnosis problem of ASD. However, more work needs to be carried out with a larger number of samples to eliminate the possible effects of misclassification and establish the method's practical validity before putting it into clinical practice.

5. Conclusion

The proposed framework successfully classified ASD children from the TD ones using their EEG signals collected over the FG brain region during EFE tasks. Linear and nonlinear measurements of IMFs were assessed using statistical analysis and a machine learning framework. A high prediction accuracy of 100% for the happy stimulus dataset was achieved. Thus, the study revealed general impairment in recognizing the positive EFE in ASD children. Besides, EEG complexity, as indexed by SampEn measure through the alpha and theta brain waves, and the linear features extracted from the delta band components may be considered biomarkers for detecting ASD children. Nevertheless, more work needs to be carried out in the future, with a larger ASD and TD children population, to bring it into clinical practices effectively. Investigating the proposed method with a sample of younger children is also required.

Acknowledgement

This work is partly funded by the scholarship programme of the University of Jeddah, Jeddah, Saudi Arabia. Dalal Bakheet is awarded this scholarship for a Ph.D. study at the University of Southampton, Southampton, UK.

References

- [1] L. Zwaigenbaum and M. Penner, "Autism spectrum disorder: Advances in diagnosis and evaluation," *BMJ*, vol. 361, pp. 1–16, 2018.
- [2] M. Ahmadlou and H. Adeli, "Electroencephalograms in diagnosis of autism," in *Comprehensive Guide to Autism* (V. Patel, V. Preedy, and C. Martin, eds.), pp. 327–343, New York, NY: Springer, 2014.
- [3] M. Black, N. Chen, K. Iyer, O. Lipp, S. Bolte, M. Falkner, T. Tan, and S. Girdler, "Mechanisms of facial emotion recognition in autism spectrum disorders: Insights from eye tracking and electroencephalography," *Neurosci Biobehav Rev.*, vol. 80, pp. 488–515, 2017.
- [4] R. Monteiro, M. Simoes, J. Andrade, and M. Branco, "Processing of facial expressions in autism: a systematic review of eeg / erp evidence," *Rev J Autism Dev Disord*, vol. 4, pp. 255–276, 2017.
- [5] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol Heart Circ Physiol.*, vol. 278, pp. 2039–2049, 2000.
- [6] N. Rehman, Y. Xia, and D. Mandic, "Application of multivariate empirical mode decomposition for seizure detection in eeg signals," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 1650–1653, 2010.
- [7] F. Apicella, F. Sicca, R. Federico, G. Campatelli, and F. Muratori, "Fusiform gyrus responses to neutral and emotional faces in children with autism spectrum disorders: a high density erp study," *Behavioural Brain Research*, vol. 251, pp. 1–8, 2013.
- [8] W. Jamal, S. Das, I.-A. Oprescu, K. Maharatna, F. Apicella, and F. Sicca, "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates," *Journal of Neural Engineering*, vol. 11, p. 046019, 2014.
- [9] S. Pincus, "Approximate entropy as a measure of system complex-

Table 6

Comparison of state-of-the-art methods employed on the EFE-based ERP studies for ASD diagnosis.

Authors	Features	Dataset	Evaluation methods	Results	Frequency analysis	Emotional expression deficit
Apicella <i>et al.</i> [7]	Peak latency and amplitude of ERP components	Dataset used in this study	ANOVA	General delayed latencies and reduced amplitudes of all ERPs in ASD children regardless of the face expressions	Did not study	None
Jamal <i>et al.</i> [8]	Complex network parameters to measure brain connectivity	Dataset used in this study	SVM	94.7% ACC, 85.7% TPR, 100% TNR using the combination of the features of all the three stimuli	Did not mention	Did not study
Khuntia <i>et al.</i> [29]	Multivariate pattern of the time points	Dataset used in this study	CPCA	Around 81% in the time domain and 84% in the frequency domain analysis	Alpha and beta bands are best to identify ASD (spectrally processed using STFT)	None
Yeung <i>et al.</i> [26]	Functional connectivity (theta coherence)	TD (n = 18) and ASD (n = 18)	ANOVA	ASD children exhibited abnormal patterns when recognizing different EFE	Theta coherence can identify ASD (they just explored theta wave using FFT)	Impairment in recognition of the happy EFE in ASD
Dominguez <i>et al.</i> [25]	Functional connectivity (imaginary part of coherency)	TD (n = 31) and ASD (n = 72)	ANOVA	ASD children exhibited enhanced synchronization during EFE	Lower frequency bands are best to identify ASD (spectral analysis method did not mention)	Did not mention
			L-SVM	80% ACC was reached using feature vectors of the imaginary part of coherency of all frequency bands		
Current study	SampEn, Min., Std. in the MEMD domain	TD (n = 12) and ASD (n = 12)	ANOVA	Significant differences were found for the happy stimulus using the features extracted from IMF5, IMF6 and IMF7	The combination of alpha, theta, and delta bands were found to best identify ASD (spectrally processed using MEMD and HT).	Impairment in recognition of the happy EFE in ASD
			DA, SVM, and kNN	100% ACC, 100% TPR, 100% TNR was reached for the happy stimulus by combining all features and extracting them from IMF5-IMF7 combination		

ity.," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[10] S. Pincus, "Assessing serial irregularity and its implications for health," *Ann N Y Acad Sci.*, vol. 954, pp. 245–267, 2001.

[11] J. Bruhn, H. Ropcke, and A. Hoefl, "Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia," *Anesthesiology*, vol. 92, no. 3, pp. 715–726, 2000.

[12] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng,

N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[13] D. Gabor, "Theory of communication. part 1: The analysis of information," *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, pp. 429–441, 1946.

[14] S. Mallat, "A theory for multiresolution signal decomposition: the

- wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [15] S. Sanei and J. Chambers, *EEG signal processing*. Chichester, England: John Wiley and Sons, 2007.
- [16] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic seizure detection in eegs using time-frequency analysis,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.
- [17] K. Vehkalhti, “The concise encyclopedia of statistics by yadolah dodge,” *International Statistical Review*, vol. 76, no. 3, pp. 460–461, 2008.
- [18] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to Pattern Recognition: A Matlab Approach*. USA: Academic Press, Inc., 2010.
- [19] MathWorks, “Statistics and machine learning toolbox,” 2021.
- [20] A. M. Molinaro, R. Simon, and R. M. Pfeiffer, “Prediction error estimation: a comparison of resampling methods,” *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005.
- [21] A. Catarino, O. Churches, S. Baron-Cohen, A. Andrade, and H. Ring, “Atypical eeg complexity in autism spectrum conditions: A multi-scale entropy analysis,” *Clinical Neurophysiology*, vol. 122, no. 12, pp. 2375–2383, 2011.
- [22] S. Khan, A. Gramfort, N. R. Shetty, M. Kitzbichler, S. Ganesan, J. Moran, S. Lee, J. Gabrieli, H. Tager-Flusberg, R. Joseph, and et al., “Local and long-range functional connectivity is reduced in concert in autism spectrum disorders,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 8, pp. 3107–3112, 2013.
- [23] M. Balconi and C. Lucchiari, “EEG correlates (event-related desynchronization) of emotional face elaboration: A temporal analysis,” *Neuroscience Letters*, vol. 392, no. 1-2, pp. 118–123, 2006.
- [24] M. Balconi and U. Pozzoli, “Arousal effect on emotional face comprehension,” *Physiology & Behavior*, vol. 97, no. 3-4, pp. 455–462, 2009.
- [25] L. G. Dominguez, J. Stieben, J. L. P. Velazquez, and S. Shanker, “The imaginary part of coherency in autism: Differences in cortical functional connectivity in preschool children,” *PLoS ONE*, vol. 8, no. 10, 2013.
- [26] M. K. Yeung, Y. M. Han, S. L. Sze, and A. S. Chan, “Altered right frontal cortical connectivity during facial emotion recognition in children with autism spectrum disorders,” *Research in Autism Spectrum Disorders*, vol. 8, no. 11, pp. 1567–1577, 2014.
- [27] B. Corden, R. Chilvers, and D. Skuse, “Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in aspergers syndrome,” *Neuropsychologia*, vol. 46, no. 1, pp. 137–147, 2008.
- [28] M. K. Yeung, T. L. Lee, and A. S. Chan, “Impaired recognition of negative facial expressions is partly related to facial perception deficits in adolescents with high-functioning autism spectrum disorder,” *Journal of Autism and Developmental Disorders*, vol. 50, no. 5, pp. 1596–1606, 2019.
- [29] A. T. Khuntia, R. Divakar, F. Apicella, F. Muratori, and K. Das, “Visual processing and attention rather than face and emotion processing play a distinct role in ASD: an EEG study,” 2019.