WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM

e-ISSN: 2788-5879 p-ISSN: 2788-5879

Lie Detection: Truth Identification from EEG Signal

Israa J. Mohammed[1,*] and Dr. Loay E. George[2]

[1]Informatics Institute for Postgraduate Studies/ Iraqi Commission for Computers & Informatics (IIPS/ICCI)
[2]University of Information Technology and Communication.

*Corresponding Author: Israa J. Mohammed

DOI: https://doi.org/10.31185/wjcm.78
Received: September 2022; Accepted: November 2022; Available online: December 2022

ABSTRACT: This study investigated the approach of extracting features from single EEG channels when the minimum number of features in Electroencephalogram (EEG) channels, hence the visibility of using sets of features extracted from a single channel. The feature sets considered in previous studies are utilized to establish a combined set of features extracted from one channel. The feature is the set of statistical moments. Publicly available EEG datasets like the Dryad dataset, obtained from 15 participants, are tested into a support vector machine classifier. The 12 channels were trained separately, where each channel was divided into a different number of blocks, and the results indicated that some channels were bad. Some were very encouraging, reaching 100% in the number of blocks 16 in channels 8 and 12. In this article, the comparison of ANN algorithm test results published in a previous article with SVM algorithm test results for the same tested features and channels will be presented.

Keywords: EEG signal processing, Statistical Moments, Support vector machine

1. INTRODUCTION

Several physiological measurements are being taken to detect a lying individual among a group of people. Heart rate [1] [2], respiratory movement, blood pressure, galvanic skin response (GSR), and functional magnetic resonance imaging are some of the measures (fMRI). These approaches are vulnerable to loss, theft, or falsification.

EEG is the measurement [3] of the brain's electrical impulses produced by neuronal activation. In reaction to external environmental happenings, the human brain generates electrical potential, known as event-related potential (ERP) [4]. Unlike traditional biometrics, electroencephalogram (EEG) signals are relatively new biological features that have lately been investigated for lie detection due to their robustness against forgery and theft [5].

Different time domain features [6] and frequency domain features [7] have been utilized to investigate various aspects of EEG signals. Similarly, classification techniques such as support vector machine (SVM) [8] and artificial neural network (ANN) [9]. There have been various proposed approaches for EEG-based lie detection. “Syed Anwar et al., 2019 [1]; have discussed employing a wearable EEG headset to detect lies using Event-Related Potential (ERP) data. Researchers explored brain activity to identify sensitive information as an alternative to the polygraph test in this study. The experiment included 14 channels and 10-20 systems in this study.

The first two principal channels produce the best findings, accounting for more than 80% of data variation. The classification method (SVM) for lie detection improves the system's accuracy using fewer EEG channels. A portable EEG recording device with a low channel commercially available EMOTIV headset is used to create an upgraded deception detection system. Frequency characteristics collected include Peak-to-Peak, Peak-to-Peak Slope, Time Window, and Amplitude and Absolute Amplitude (AAMP) (deleted the zeros). ENT and A.P., as well as several other metrics, benefit advantage vectors. The testing set was created to be as realistic as possible by using a group of kids close to each other.

The system has an 83% accuracy rate. Changing the cost parameters of the SVM with a specific training sequence can give additional benefits.

"D.H. Yohan Kulasinghe, 2019 [2], has discussed using EEG technology and machine learning to detect lying. Machine learning methods that may analyze EEG data include SVM, k-Means, ANN, and Linear Classifiers. The scalp EEG is captured using the 10-20 system technology. When someone discusses a mixture of truth and lies, it may take some time, which might be beneficial. To assess EEG signals, signal amplitude, wavelength, frequency, and voltage were employed as classification model characteristics. It detected dishonesty using a frontal pole (Fp) and temporal region (T) cues. Because those areas are responsible for logical thinking, reasoning, judgment-related processes, emotional reactions, and remembering. The Fast Fourier Transform (FFT) method converts complex EEG waveforms into simple waveforms. Use the algorithm to distinguish between the truth and the falsehood. SVM performs admirably on a single data point. The approach's accuracy is 86%, according to the most convenient feature for detecting lies".

Yijun Xiong et al., 2020 [3] compared the EEGs of the two groups during lie detection (L.D.) trials using the chaotic phase synchronization (P.S.) technique. Twenty participants' EEGs were recorded in the L.D. study using a three-stimulus approach. P.S. employed the statistical metric known as the Phase Locking Value (PLV) for a few stimuli in the L.D. investigation. A strong and larger PLV was observed in the guilty group compared to the innocent group, indicating a clear spatial and temporal difference in P.S. It was utilized to examine the interconnectedness of their frontal, temporal, central, and parietal regions to identify their tricks. The researchers did this by coordinating phase synchronization patterns across 12 EEG channels. Analyzing P.S. using the EEG data from a limited-detection L.D. experiment, we investigated the P.S. between various EEG activities recorded from different brain regions. Ten healthy college students (9 males, mean age of 22.3 years) were recruited for the study (22.3 is the average). Three unique types of stimulation were used in this method. The electrodes were arranged using the 10-20 system. In total, 14 channels were used to record horizontal and vertical EOGs. Using PLV-based characteristics, an SVM was constructed to distinguish between truthful and dishonest mental states.

However, in order to be practical, quick, and accurate, an EEG-based lie detection system needs to go through a few straightforward stages. The acquisition process should be easy and simple to avoid disturbing the user. Therefore, to reduce the system's complexity while maintaining high system accuracy, the least number of electrodes (or channels) must be linked to the user's scalp, and the minimal number of features must be checked in each channel.

In previous work, [4] discussed a method for extracting features from individual EEG channels using a minimum number of features to keep the detection system's complexity to a minimum; features were evaluated in ANN, and competitive results were obtained.

"In this work, the same approach to extracting features from EEG channels is discussed, using the same feature types proposed in previous works. This paper is organized as follows: Section 2 presents the description of used datasets and the proposed methods, Section 3 presents the description ofclassification , Section 4 discusses the experiments result, Section 5 discusses previous works related to this paper, and Section 6 presents conclusions".

2. METHODOLOGY

This study uses electroencephalogram (EEG) signals as a set of statistical features to build a lie detection system. Two approaches are adopted; the first is to extract features from a single channel (or electrode), and the second is to use the least number of features extracted from the channel. EEG-based automatic lie detection and truth-identification system apply four main stages (i.e., preprocessing, feature extraction, feature selection, and classification phase) to the input EEG. The preprocessing includes two steps, normalization, and framing. Feature extraction contains two sets of features; Spectral features and statistical moments. In feature selection, including selecting the most discriminative features. The SVM classifier is used for the classification task.

2.1 DATASETS

The publicly accessible EEG Dryad dataset [4] assesses the system's performance on lie detection tasks. It is a public, free dataset that is fairly huge. It is made up of 12 EEG channels, each with 16384 samples.

2.2 PROPOSED SYSTEM

The proposed methods in [4], which worked under the approach (using a single channel and a minimum number of features), are tested in ANN in this study under the same approach (is tested in SVM).

2.2.1. Proposed Features

Two sets of features were used for the lie detection system in the previous study; these are the Prosodic Features and statistical moments features:

2.2.2. Feature Selection

In this stage, several different feature combinations have been evaluated and compared to identify those that provide detection rates that are the highest attainable and are discussed In detail [4].

• **Prosodic Features**

A type of measurement feature concerned with timing, articulation, duration, and zero crossing (ZCR) is the rate at which a signal changes from positive to zero to negative (ZCR) is the other way around [5] [6] used in [4]

$$ZCR_{(n)} = \frac{1}{2\,(L-1)} \sum_{i=1}^{L-1} |sgn\,(x_{i+1}) - sgn\,(x_i)| \tag{1}$$

Where S is a signal of length and 1R<0 is an indicator function (is a function that maps elements of the subset to one and all other elements to zero).

• **Statistical Moments Feature**

Moments refer to how much a given quantity differs from its mean or any pivot point in terms of mass, force, histogram intensity, frequency transform coefficients, and other kinds of coefficients with certain geometrical distributions [5] [7].

Moments can be classified into many categories. Mathematically, moment features for a frame are calculated to characterize its behavior and extract critical features. These features are described by (1), (2),(3),(4),(5), and (6) [8] [9] [10] [11]:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2}$$

Where μ *is mean* , N denotes the total number of samples, and x_i is the sample.

$$\sigma = \sqrt{\frac{1}{N} \sum\nolimits_{i=1}^{N} (x_i - \mu)^2} \tag{3}$$

Where σ is the standard deviation, N is the total number of samples, x_i is the sample, μ is the mean.

$$Skew = \frac{\sum_{i=1}^{N} (x_i - \mu)^3}{N\,\sigma^3} \tag{4}$$

$$Kurt = \frac{\sum_{i=1}^{N} (x_i - \mu)^4}{N\,\sigma^4} \tag{5}$$

Where N is the total number of samples, xi is the sample, μ is the mean, and σ is the standard deviation.

$$\mu_5 = \frac{1}{N} \sum\nolimits_{i=1}^{N} (x_i - \mu)^5 \tag{6}$$

Where N is the total number of samples, xi is the sample, μ is the mean.

$$p = \frac{1}{N} \sum\nolimits_{i=1}^{N} (|x_i|)^2 \tag{7}$$

Where P is Signal Power, N is the total number of samples, xi is the sample.

All the features mentioned above have been used in the published article [4].

2.2.3. Feature Selection

In this stage, several different feature combinations have been evaluated and compared to identify those that provide detection rates that are the highest attainable and are discussed In detail [4].

3. CLASSIFICATION

Classification is the process of categorizing feature extraction findings based on mental activities or supplied inputs. To obtain the desired EEG signal, it must correspond to mental tasks and motivations. The Support Vector Machine (SVM) is a popular classification technique because of its accurate results and short computing time. SVM is a classification and regression prediction algorithm [12]. SVM and ANN are both class-supervised learning methods that are comparable. SVM has been around for a long time, mixing computational techniques like margin hyperplane and kernel [10].

In the present work, the fine gaussian kernel is used with SVM. Following feature extraction, initially, data are divided into 50% and 50%. Testing data are taken as 50%. Out of the remaining 50%, training data are selected. MATLAB is used to train an SVM algorithm for the classification of detections.

SVM “is a collection of relevant supervised machine learning techniques that analyses data and identify data structures for categorization” [13]. SVM-based classification has been shown to strike the ideal balance between accuracy gained on a limited stock of training data and generalization on test data [14].

In other words, SVM “is a technique to obtain the most probable hyperplane to separate two classes” [15]. “It is done by measuring the hyperplane’s margin and determining its maximum point. Margin is defined as the distance between the corresponding hyperplane and the nearest pattern from each class. Moreover, this nearest pattern is called a support vector” [16].

4. EXPERIMENTAL RESULTS

This paper presents and discusses the findings of a few tests run to assess the established system’s performance. The Microsoft Visual Studio 2012’s C# programming language and MATLAB were used.

Dryad datasets were used to test the proposed system’s accuracy with all the proposed feature extraction methods. Dryad datasets are relatively large (12 EEG channels and two EOG channels), where each set of features is extracted from a single channel. The best attained system recognition rate was 100% for some feature sets and channels for each proposed feature extraction method using all the datasets.

The training and testing results of the algorithm SVM will be displayed and compared with the results of the algorithm ANN

This table shows the training results of some channels in the system with dataset samples tested in the SVM algorithm; it gave the best result in blocks number 16:

Table 1. The result of the train and test features in blocks number is 16 in channel 1

Channel number	Channel-name	Block Num	Feature number	Features set	Accu-racy
1	AF3	16	7	Mean, Std, Skew, Kurt, Power, Zero, Pow5	81.2%
1	AF3	16	1	Mean	75%
1	AF3	16	1	Std	62.5%
1	AF3	16	1	Skew	56.2%
1	AF3	16	1	Kurt	62.5%
1	AF3	16	1	Power	68.8%
1	AF3	16	1	Zero	68.8%
1	AF3	16	1	Pow5	62..5%
1	AF3	16	3	Mean, Power, Zero	81.2%
1	AF3	16	2	Mean, Power	75%

In general, the results presented in the above tables show that there are 12 channels tested separately. The test results indicate that channels 8 and 12 were the best, as their test results reached 100%, and these results are considered encouraging for using one channel.

(7), (8), (9) (10) will show some of the best results of trained and tested data in the ANN algorithm discussed in the published article [4] to compare with the above results mentioned above.

There are types in the kernel function, and several types were tried, as in the following table, and the best results were observed in the kind of function kernel in Fine Gaussian SVM.

When some of the results of the test of the algorithm SVM and the algorithm ANN were presented, it was concluded that the results of the algorithm SVM are good in the number of blocks 16. The strongest channels were 8 and 12 when compared with the results of the ANN; the results were stronger and more encouraging than the results of the SVM algorithm. The best algorithm with the data set used is the ANN algorithm.

Table 2. The result of the train and test features in blocks number is 16 in channel 4

Channel number	Channelname	Block Num	Feature number	Features set	Accuracy
4	FC3	16	7	**Mean, Std, Skew, Kurt, Power, Zero, Pow5**	87.5%
4	FC3	16	1	Mean	81.2%
4	FC3	16	1	Std	56.2%
4	FC3	16	1	Skew	50%
4	FC3	16	1	Kurt	56.2%
4	FC3	16	1	power	75%
4	FC3	16	1	Zero	62.5%
4	FC3	16	1	Pow5	56.2%
4	FC3	16	3	Mean, power, Zero	93.8%
4	FC3	16	2	Mean, power	75%

Table 3. The result of the train and test features in blocks number is 16 in channel 8

Channel number	Channelname	Block Num	Feature number	Features set	
8	O2	16	7	Mean, Std, Skew, Kurt, Power, Zero, Pow5	93.8%
8	O2	16	1	Mean	100%
8	O2	16	1	Std	81.2%
8	O2	16	1	Skew	87.5%
8	O2	16	1	Kurt	75%
8	O2	16	1	power	93.8%
8	O2	16	1	Zero	68.8%
8	O2	16	1	Pow5	87.5%
8	O2	16	3	Mean, Skew, power	100%
8	O2	16	2	Mean, power	93.8%

Table 4. The result of the train and test features in blocks number is 16 in channel 9

Channel number	Channelname	Block Num	Feature number	Features set	Accuracy
9	P8	16	7	Mean, Std, Skew, Kurt, Power, Zero, Pow5	93.8%
9	P8	16	1	Mean	87.5%
9	P8	16	1	Std	75%
9	P8	16	1	Skew	68.8%
9	P8	16	1	Kurt	68.8%
9	P8	16	1	power	100
9	P8	16	1	Zero	62.5%
9	P8	16	1	Pow5	81.2%
9	P8	16	3	Mean, power, Pow5	100%
9	P8	16	2	Mean, pow5	93.8%

Table 5. The result of the train and test features in blocks number is 16 in channel 11

Channel number	Channelname	Block Num	Feature number	Features set	Accuracy
11	FC6	16	7	Mean, Std, Skew, Kurt, Power, Zero, Pow5	93.8%
11	FC6	16	1	Mean	93.8%
11	FC6	16	1	Std	81.2%
11	FC6	16	1	Skew	75%
11	FC6	16	1	Kurt	81.2%
11	FC6	16	1	power	93.8%
11	FC6	16	1	Zero	100
11	FC6	16	1	Pow5	81.2%
11	FC6	16	3	Mean, power, Zero	93.8%
11	FC6	16	2	power, Zero	93.8%

Table 6. The result of the train and test features in blocks number is 16 in channel 12

Channel number	Channelname	Block Num	Feature number	Features set	Accuracy
12	F9	16	7	Mean, Std, Skew, Kurt, Power, Zero, Pow5	100%
12	F9	16	1	Mean	100%
12	F9	16	1	Std	81.2%
12	F9	16	1	Skew	93.8%
12	F9	16	1	Kurt	87.5%
12	F9	16	1	power	93.8%
12	F9	16	1	Zero	75%
12	F9	16	1	Pow5	81.2%
12	F9	16	3	Mean, Skew, power	100%
12	F9	16	2	Mean, power	93.8%

Table 7. The best-attained results of training and testing samples using one feature with a blocks number are 16

Channel number	Channelname	Block number	Features number	Features set	Successful samples	Failed samples	Accuracy
11	FC6	16	1	Zero crossing	32	0	100%
11	FC6	16	1	Power	31	1	96.9%
8	O2	16	1	Mean	31	1	96.9%
9	P8	16	1	Mean& power	31	1	96.9%
12	F9	16	1	Mean	31	1	96.9%
8	O2	16	1	Power	30	2	93.8%
12	F9	16	1	Std. Dev.	30	2	93.8%

Table 8. The best-attained results of training and testing sets using two features with blocks number are 16

Channel number	Channelname	Block number	Features number	Features set	Successful samples	Failed samples	Accuracy
11	FC6	16	2	Power, Zero Crossing	32	0	100%
12	F9	16	2	Mean, Std	32	0	100%
8	O2	16	2	Mean, power	31	1	96.9%
9	P8	16	2	Mean, Power	31	1	96.9%
10	T8	16	2	Mean, Power	29	3	90.6%

Table 9. **Thebest-attained results of training and testing samples using three features with a blocks number are 16**

Channel number	Channelname	Block number	Features number	Features set	Successful samples	Failed samples	Accuracy
11	FC6	16	3	Mean, Power, Zero Crossing	32	0	100%
4	FC3	16	3	Mean, power, Zero crossing	29	3	90.6%

Table 10. **The best-attained results of training and testing samples using four features with blocks number16**

Channel number	Chan-nelname	Block number	Features number	Features set	Success. samples	Failed samples	Accu-racy
12	F9	16	4	Mean, StdDev, Skew, Power	32	0	100%
8	O2	16	4	Mean, power, Kurtosis, Pow5	31	1	96.9%
9	P8	16	4	Mean, Power, Kurtosis, Pow5	31	1	96.9%
10	T8	16	4	Mean, Skew, Power, pow5	30	2	93.8%

Table 11. **The best results of training and testing data in several types of kernel functions in SVM**

Channel number	Chan-nel name	Block size	Features number	Accuracy Fine Gaussian SVM	Accuracy Linear SVM	Accuracy Quadratic SVM	Accuracy Cubic SVM
8	O2	16	**7**	93.8%	93.8%	93.8%	93.8%
8	O2	16	1	93.8%	87.5%	87.5%	93.8%
8	O2	16	3	100%	93.8%	100%	87.5%
8	O2	16	1	87.5%	75%	81.2%	75%
12	F9	16	7	100%	93.8%	93.8%	93.8%
12	F9	16	1	100%	87.5%	93.8%	93.8%
12	F9	16	2	93.8%	93.8%	93.8%	93.8%
12	F9	16	3	100%	100%	100%	100%

5. COMPARISON WITH RELATED WORKS

Many published studies on EEG-based lie detection systems have shown promising results, although many utilized more than one channel or feature to detect the deceptions. Table (12) shows that this paper's results are comparable to those of other papers published on the SVM algorithm.

Table 12. **Comparisons of the SVM algorithm are based on the number of channels and features employed.**

Authors	No. of Channel	No. of Features	Accu-racy
Syed Anwar et al., [1]	12 channels	five discriminative features are used	83%
D.H. Yohan Kulasinghe, [2]	2 Channels	four discriminative features are used	86%
Yijun Xiong et al., [3]	12 channels	Three groups of features	—-
Proposed Work	Single Channel	Only one or merge (2 or 3 or 4) and all 7 Features test in some cases give	100%

6. CONCLUSIONS AND FUTURE WORK

In this paper, an approach to extracting features from user EEG signals is adopted; the features proposed in previous studies are tested to check the discriminative degree of these features when tested in the SVM algorithm.

This approach has good performance in the lie detection system, but the performance of the ANN algorithm for most types of features is better than the SVM approach. This approach also keeps the computational complexity low and uses a single channel. After completing

This study showed that one or two EEG channels are enough to extract discriminate features and detect the lies when the proposed method was tested on the available datasets. A new type of statistical moment is recommended as a new feature for the EEG-lie detection system and can be tested on other data sets.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] S. Anwar, T. Batool, and M. Majid, "Event Related Potential (ERP) based Lie Detection using a Wearable EEG headset," *Proc. 2019 16th Int. Bhurban Conf. Appl. Sci. Technol. IBCAST 2019*, pp. 543–547, 2019.

[2] D. H. Y. Kulasinghe, "Faculty of information technology university of Moratuwa - Sri Lanka," in Using EEG and Machine Learning to perform Lie Detection."

[3] Y. Xiong, L. Gu, and J. Gao, "Phase synchrony and its application to lie detection," *Proc. 2020 IEEE Int. Conf. Power, Intell. Comput. Syst. ICPICS 2020*, pp. 726–729, 2020.

[4] I. J. Mohammed and L. E. George, "Lie Detection and Truth Identification form EEG signals by using Frequency and Time Features," *J. Algebr. Stat*, vol. 13, no. 3, pp. 4102–4121, 2022.

[5] D. Dacunha-Castelle and M. Duflo *Probability and Statistics*, vol. II, 2012.

[6] N. Zhou and L. Wang, "A modified T-test feature selection method and its application on the HapMap genotype data," *Genomics. Proteomics Bioinformatics*, vol. 5, no. 3-4, pp. 242–249, 2007.

[7] S. Dodia, D. R. Edla, A. Bablani, and R. Cheruku, "Lie detection using extreme learning machine: A concealed information test based on short-time Fourier transform and binary bat optimization using a novel fitness function," *Comput. Intell*, vol. 36, no. 2, pp. 637–658, 2020.

[8] B. Liu, "Uncertain risk analysis and uncertain reliability analysis," *J. Uncertain Syst*, vol. 4, no. 3, pp. 163–170, 2010.

[9] S. Dey, B. Al-Zahrani, and S. Basloom, "Dagum distribution: Properties and different methods of estimation," *Int. J. Stat. Probab*, vol. 6, no. 2, pp. 74–92, 2017.

[10] K. P. Balanda and H. L. Macgillivray *Kurtosis : A Critical Review*, vol. 42, pp. 111–119, 2012.

[11] J. Cohen, "Statistical power analysis," *Curr. Dir. Psychol. Sci*, vol. 1, no. 3, pp. 98–101, 1992.

[12] I. E. Naqa and M. J. Murphy *What is machine learning?," in machine learning in radiation oncology*, pp. 3–11, 2015. Springer.

[13] S. B. Maind and P. Wankar, “Research paper on basic of artificial neural network,” *Int. J. Recent Innov. Trends Comput. Commun*, vol. 2, no. 1, pp. 96–100, 2014.
[14] I. Steinwart and A. Christmann, “Support vector machines,” 2008. Springer Science & Business Media.
[15] A. Bablani, D. R. Edla, V. Kupilli, and R. Dharavath, “Lie Detection Using Fuzzy Ensemble Approach with Novel Defuzzification Method for Classification of EEG Signals,” *IEEE Trans. Instrum. Meas*, vol. 70, pp. 2021–2021.
[16] Y. Kulasinghe and D. H. Y. Kulasinghe, “Using EEG and Machine Learning to perform Lie Detection Kinect sensor for skeleton abnormality detection View project Software Effort Estimation Model Based on Story Points for Agile Development View project Using EEG and Machine Learning to perform Lie D.” https://www.researchgate.net/publication/335095404.