EEG-Based Lie Detection using Machine Learning

Biometric Systems Project Report
`https://github.com/olja-coro/lie_detection_eeg`

Davide Perniconi - 1889270
Olja Corovencova - 2249558
Leonardo Lavezzari - 1984079

Abstract

In this project, we investigate the use of machine learning approaches for EEG-based lie detection. Inspired by recent entropy-based methods, we evaluate subject-dependent and cross-subject classification pipelines, conducting channel-wise analysis on the LieWaves dataset. We first successfully reproduce the results reported in the reference study, thereby validating the robustness of the proposed entropy-based feature extraction. In addition to classical machine learning algorithms, we explore deep learning approaches to assess whether potentially more expressive models can further enhance performance. Specifically, we evaluate a Multilayer Perceptron (MLP) trained on the same entropy-based features, as well as a sequential architecture trained directly on raw recorded signals. Building upon this foundation, we extend the original work by performing a more comprehensive analysis of multi-channel EEG configurations. In particular, we systematically evaluate all 32 possible combinations of the 5 available EEG channels, providing a detailed investigation of the impact of channel fusion on classification performance. For single-channel experiments, the most informative channels were found to be Pz, T8 and T7, consistent with prior studies. However, in contrast to the original work, our multi-channel analysis revealed that the best-performing combinations were not simply those formed by aggregating the top single channels. Interestingly, channels that were less informative in isolation were found to contribute substantially when combined with other electrodes, highlighting the complementary nature of multi-channel EEG information for lie detection.

1 Introduction

Lie detection is an important research area with applications in forensic investigations, security screening and psychological assessment. Traditional lie detection techniques, such as polygraph tests, rely on peripheral physiological responses such as heart rate, respiration, and skin conductance [1]. These techniques are known to be sensitive to countermeasures and are considered unreliable as they indirectly measure deceptive behaviour rather than its neural origin [2]. In recent years, neurophysiological approaches have gained increasing attention as deception is fundamentally a cognitive process involving memory, attention, and decision-making.

Electroencephalography (EEG) is a particularly promising method for this purpose as it provides a low-cost and portable means of recording brain activity with high temporal resolution. EEG signals can capture the rapid neural dynamics associated with the cognitive processes that occur during truthful and deceptive responses, making them suitable for real-time lie detection applications. However, EEG-based lie detection remains challenging due to the noisy nature of the signals and the difficulty of extracting discriminative features from brain activity. For this reason, machine learning techniques have become central to modern EEG analysis, demonstrating promising performance in distinguishing between truthful and deceptive conditions. By learning patterns directly from data, these methods can automatically model complex, nonlinear relationships between EEG features and cognitive states.

The present project draws inspiration from the recent multi-scale entropy-based approach proposed by Li *et al.* [3] The objective is to assess the efficacy of classical machine learning algorithms by reproducing their experiments from scratch to confirm their results. The subsequent stage of the investigation is to determine whether deep learning approaches, specifically a shallow multi-layer perceptron and a recurrent architecture, can achieve superior outcomes.

The present study evaluates both subject-dependent and cross-subject classification pipelines on the LieWaves dataset [4], and it also provides insights into the role of EEG channels in distinguishing truthful and deceptive cognitive states.

2 Dataset Description

The experiments conducted in this project are based on the LieWaves dataset, a publicly available EEG dataset specifically designed for deception detection research. The dataset contains EEG recordings collected from 27 subjects using a wearable and portable EEG device, the Emotiv Insight headset, which provides 5 channels: AF3, T7, Pz, T8 and AF4. Two experimental sessions were carried out, with each subject taking on the roles of truth-teller and deceiver in separate ones. During each session, subjects were presented with visual stimuli consisting of images of beads placed in front of them. The experimental protocol followed a repetitive pattern, where a black screen was displayed for a few seconds, followed by bead images, and then another black screen. In accordance with the designated role, subjects were instructed to respond truthfully or deceptively by pressing designated buttons. A total of 75 seconds of EEG data per subject and per session were collected. In addition to raw signals, the dataset provides different preprocessed versions of the time-series but only band-pass filtered data (in the range 0.5-45 Hz) was used in our experiments, as suggested by Li *et al.*

3 Methodology

This section presents the methodology employed in this work. It first outlines the feature extraction process used to derive multi-scale entropy based representations from EEG time-series data. Subsequently, the classification models considered in the study, including both classical machine learning and deep learning approaches, are described. Finally, the training and evaluation protocols adopted to assess model performance under subject-dependent and cross-subject settings are detailed.

3.1 Feature extraction

Time-series data was converted into multi-scale entropy based representations following the feature extraction strategy described in the reference paper. The goal is to capture EEG signal complexity across single-scale, multi-scale temporal and multi-band spectral perspectives using three complementary entropy descriptors:

- **Fuzzy entropy** (FE): provides a single scalar quantifying the regularity/complexity of the signal. The $\phi^m(r,n)$ term in Equation 1 can be interpreted as the average probability that two m-dimensional patterns randomly drawn from the original signals match each other.

$$\mathrm{FE}(m,r,n) = \ln \phi^m(r,n) - \ln \phi^{m+1}(r,n) \tag{1}$$

- **Time-shifted multi-scale fuzzy entropy** (TSMFE): k time-shifted subsequences are constructed across different time scales: $y_j^{(k)} = \{x_j, x_{j+k}, x_{j+2k}, \ldots\}$ with $j = 1, 2, \ldots$, and TSMFE(k) is defined as the average of the FE values of all k time-shifted subsequences, as shown by the following equation:

$$\mathrm{TSMFE}(k) = \frac{1}{k}\sum_{j=1}^{k} \mathrm{FE}(y_j^{(k)}, m, r, n). \tag{2}$$

  The final TSMFE feature vector is

$$\mathrm{TSMFE} = [\mathrm{TSMFE}(1), \mathrm{TSMFE}(2), \ldots, \mathrm{TSMFE}(k_{max})].$$

- **Hierarchical multi-band fuzzy entropy** (HMFE): decomposes the signal into frequency sub-bands via discrete wavelet transform (DWT) and computes FE for each reconstructed band, generating a compact spectral-complexity representation.

$$\mathrm{DWT}_L(x) = \{A_L, D_L, D_{L-1}, \ldots, D_1\}$$
$$\mathrm{HMFE} = [\mathrm{FE}_{AL}, \mathrm{FE}_{DL}, \mathrm{FE}_{D_{L-1}}, \ldots, \mathrm{FE}_{D_1}].$$

For this phase of feature extraction, the following parameter values were chosen (in according to Li *et al.* study):

- The embedding dimension was fixed to $m = 2$. It defines the length of the pattern vectors.
- The fuzzy exponent was set to $n = 2$, which controls the shape of the fuzzy membership function and provides a smooth transition between similar and dissimilar patterns.
- The tolerance parameter was defined as $r = 0.2 \times \sigma$, where σ denotes the standard deviation of the analyzed channel raw signal, ensuring scale invariant similarity assessment.
- The length of the TSMFE vector was set to $k_{\max} = 10$, allowing the analysis of the signal complexity across multiple temporal resolutions.
- The discrete wavelet transform (DWT) was performed using the Daubechies 4 wavelet (db4) with a decomposition *level* of 4. This configuration yields one approximation and four detail components, resulting in a HMFE feature vector of length 5.

As a result, each 384 long time-series, corresponding to a visual stimuli (image of bead), was mapped to a 16-element feature vector, thereby achieving a dimensionality reduction of 95.83%: 1 (FE) + 10 (TSMFE scales) + 5 (HMFE bands).

$$\texttt{Final feature vector} = [\mathrm{TSMFE}_1, \ldots, \mathrm{TSMFE}_{k_{max}}, \mathrm{HMFE}_1, \ldots, \mathrm{HMFE}_{level+1}, \mathrm{FE}].$$

The final extracted entropy-based features are stored in a structured tensor format to support different evaluation protocols:

- `X` with shape (27, 2, 25, 5, 16), corresponding to subject, session, trial, channel, features.
- `y` with shape (27, 2, 25) and values 1 or 0 corresponding to truth or lie.

In addition, the original raw data was also stored to support experiments that require time series input (e.g., training of sequential models).

3.2 Classification Models

To reproduce the training pipeline presented in the reference paper, a set of classical machine learning classifiers was first employed using the standardized extracted entropy-based features. In addition, deep learning approaches were explored to investigate whether they could further improve deception detection performance.

3.2.1 Reference paper classifiers

The following classical machine learning algorithms were adopted:

- **Support Vector Machine with linear kernel (SVM-Linear)**: an hyperplane in the feature space is constructed to separate classes with maximal margin.
- **Support Vector Machine with radial basis function kernel (SVM-RBF)**: the RBF kernel allows the SVM to capture non-linear relationships by mapping data into a higher dimensional space.
- **k-Nearest Neighbors with $k = 3$ (kNN-3) and with $k = 5$ (kNN-5)**: the classification of a sample is based on the majority class among its k nearest neighbors in the feature space.
- **Gaussian Naive Bayes (NB)**: a probabilistic classifier that assumes feature independence and models each feature using a Gaussian distribution; classification is based on selecting the class with the highest posterior probability given the observed features.
- **Linear Discriminant Analysis (LDA)**: a linear classifier that projects the data onto a lower-dimensional space to maximize class separability; classification is performed by assigning a sample to the class whose projected mean is closest in this space.

All classifiers were used with standard parameter settings of the `Scikit-Learn` library and were trained to reproduce and confirm the reference paper results.

3.2.2 Deep learning extension

The two deep learning models were designed with `Pytorch` library and hypertuning was conducted with `Optuna` library.

- **Multi layer perceptron**: the network architecture consists of two fully connected hidden layers with ReLU activations and dropout regularization, followed by a single output neuron producing a binary classification score. This design allows the model to learn nonlinear combinations of the entropy features, standardized with `RobustScaler` class of `Scikit-Learn`, while maintaining a relatively small number of parameters, which is suitable for the limited size of the dataset. The MLP was treated as a feature-based classifier and thus directly comparable to the classical machine learning models used in the reference paper, differing only in the learning capacity of the decision function.
- **Sequential model**: it first implements a Temporal Convolutional Network (TCN) comprising 1D convolutions, which process all EEG channels collectively. This approach enables the modelling of inter-channel relationships from the input stage. A bidirectional Long-Short Term Memory (BiLSTM) then captures long-term temporal dependencies in the learned feature sequence by integrating past and future contextual information. Finally, a linear classification layer maps the final LSTM hidden state to a scalar output (logit).

3.3 Training and evaluation protocol

We reproduced the evaluation methodology adopted in the Li *et al.* work to ensure a fair comparison across classifiers and channels. All experiments were conducted at the trial level (stimuli associated with image of bead) and performance was reported in terms of classification accuracy.

- **Subject-dependent evaluation**: for the subject-dependent analysis, models were trained and evaluated independently for each partecipant, following a Leave-One-Out Cross-Validation (LOOCV) scheme. For a given subject, all trials from both sessions were pooled, resulting in 50 trials per subject, and at each LOOCV iteration, one trial was held out for testing, while the remaining ones were used for training. This process was repeated until each trial had served once as the test sample, and the final accuracy was computed as the average across all folds. Consistent with the reference paper, all classifiers were trained using features extracted for each subject and channel pair. This channel-wise evaluation enabled the identification of electrodes that are most informative for lie detection within individual partecipants.

- **Cross-subject evaluation**: to assess generalization across subjects, a Leave-One-Subject-Out (LOSO) evaluation protocol was adopted. In this setting, data points from one subject were completely excluded from training and used only for testing, while data from the remaining subjects were used to train the model. This procedure was repeated until each subject had served once as the held-out test subject. As in the subject-dependent analysis, models were trained separately for each EEG channel, ensuring that cross-subject performance could be directly compared across channels. This evaluation protocol reflects a realistic deployment scenario in which a model is applied to unseen subjects and is therefore more challenging than the subject-dependent case.
- **Channel combination experiments**: in order to explore the potential of combining multiple EEG channels to enhance classification performance, a series of experiments were conducted, expanding upon the initial single-channel analysis. While the original authors limited their analysis to aggregating only the top individual channels, all 32 possible combinations of the five available sensors were systematically evaluated, with features from each combination concatenated into a single feature vector per trial.

4 Results

This section reports the experimental results obtained from evaluating machine learning classifiers (ML), a multilayer perceptron (MLP) and a sequential LSTM-based model under both subject-dependent and cross-subject protocols.

Results are presented at three levels: single-channel ML classifiers, single-channel neural networks and multi-channel analysis. Accuracies are reported as percentages, along with their mean and standard deviation across partecipants. For convenient presentation of the results, 'C1' to 'C6' sequentially refer to the six types of classifiers evaluated, where C1 is for SVM-Linear, C2 is for SVM-RBF, C3 is for kNN-3,C4 is for kNN-5, C5 is for NB, and C6 is for LDA. 'S1' to 'S27' denote the 27 subjects in the dataset.

4.1 Single channel ML classifiers

The subject-dependent results, summarized in Table 1, show that the overall performance achieved in this project is comparable to, and in some cases slightly higher than, the results reported in the original paper. Across classifiers, the average LOOCV accuracies are consistently above chance level, with mean values typically ranging between approximately 80% and 85%, depending on the classifier.

A key observation is that SVM with a linear kernel (C1) achieves the highest average performance in the subject-dependent setting. This differs slightly from the reference paper, where LDA (C6) was reported as the best-performing classifier on average. Despite this difference in ranking, the absolute accuracy values remain very close, indicating that the entropy-based features extracted in this project preserve the discriminative information identified in the original work.

A notable characteristic of the subject-dependent results is the large standard deviation observed across subjects. While some subjects achieve near-ceiling accuracies, others remain close to chance level. This behavior mirrors the paper's findings and highlights that lie detection performance is highly dependent on individual neurophysiological patterns, which vary across users rather than being uniform.

The cross-subject results are reported in Table 2, showing the accuracy of the best channel for each subject and classifier. As expected, performance decreases substantially compared to the previous setting, with mean accuracies around 60–62% and high standard deviations. These results once more underscores the challenge inherent in generalizing deception-related EEG patterns across individuals.

In alignment with the reference paper, the SVM with RBF kernel (C2) achieves the best average performance in the cross-subject scenario.

Considering both evaluation protocols together, clear differences in channel behavior emerge. As shown in Figure 1, the T8 channel consistently achieves the highest average accuracy in the subject-dependent setting. The Pz channel consistently ranks second, while also exhibiting lower variance across subjects compared to T8.

In contrast, in the cross-subject evaluation, performance shifts noticeably. As reported in Figure 2, the accuracy of T8 drops substantially, whereas Pz becomes the best-performing channel overall, achieving the highest mean LOSOCV accuracy across classifiers. These observations closely align with the findings of the original paper, which identified Pz as the most representative channel for cross-subject analysis, while temporal channels (T7/T8) were more dominant in subject-dependent scenarios.

Subject	C1	C2	C3	C4	C5	C6
S01	78	80	76	80	72	74
S02	100	96	96	94	98	100
S03	72	72	66	68	76	70
S04	90	84	88	90	80	88
S05	72	74	68	72	74	70
S06	80	66	62	62	68	72
S07	86	88	80	88	86	78
S08	72	76	68	72	78	68
S09	100	100	100	100	100	100
S10	86	80	78	78	78	80
S11	84	74	74	74	66	78
S12	98	96	96	96	96	92
S13	100	96	96	94	98	100
S14	76	76	78	76	78	80
S15	86	86	72	74	82	86
S16	78	74	64	70	64	80
S17	76	60	62	58	62	72
S18	92	94	88	92	94	94
S19	74	62	58	62	58	68
S20	70	80	78	78	80	66
S21	90	90	86	84	84	82
S22	100	90	88	88	90	88
S23	74	76	68	68	60	62
S24	76	68	68	70	72	72
S25	100	96	88	90	98	100
S26	90	92	96	94	94	90
S27	90	92	92	88	88	94
Mean	**84.81**	82.15	79.04	80.00	80.52	81.63
SD	10.16	11.19	12.36	11.61	12.51	11.51

Table 1: Best single-channel LOOCV Accuracy (%) for each subject per ML Classifier

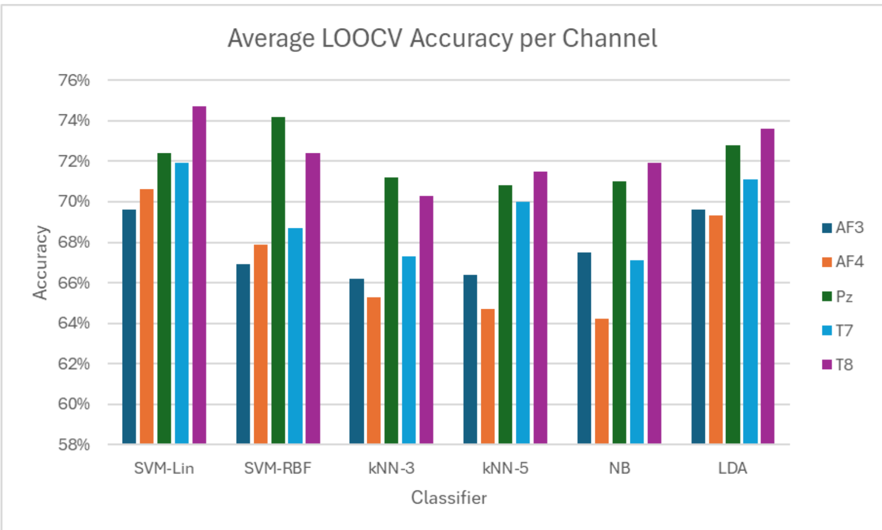

Figure 1: Average, across subject, LOOCV accuracy per channel for ML classifiers

Subject	C1	C2	C3	C4	C5	C6
S01	66	60	52	58	64	52
S02	58	58	58	48	54	54
S03	60	58	60	56	58	56
S04	68	68	66	58	76	62
S05	58	66	58	64	54	54
S06	56	54	64	70	60	60
S07	70	76	58	60	50	58
S08	58	60	60	62	54	54
S09	34	34	56	54	48	40
S10	64	72	66	60	60	66
S11	52	56	62	68	56	62
S12	66	72	68	64	96	64
S13	74	64	66	68	56	72
S14	54	80	58	58	64	54
S15	58	54	56	58	54	58
S16	66	72	46	54	56	68
S17	66	58	68	66	58	64
S18	66	70	56	64	74	66
S19	54	60	62	64	48	56
S20	76	74	60	58	80	72
S21	70	58	60	56	66	62
S22	72	62	62	62	78	66
S23	56	66	62	62	58	54
S24	56	56	58	62	52	58
S25	54	68	68	60	60	62
S26	58	58	54	58	60	54
S27	60	50	62	64	52	60
Mean	61.11	**62.37**	60.22	60.59	60.96	59.56
SD	8.47	9.33	5.09	4.80	10.95	6.74

Table 2: Best single-channel LOSOCV accuracy (%) for each left-out subject per ML classifier

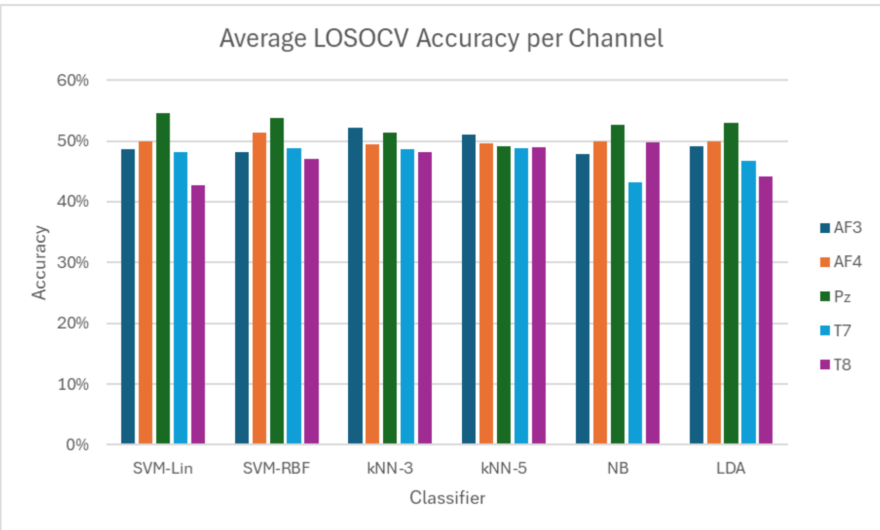

Figure 2: Average LOSOCV accuracy per channel for ML Classifiers

4.2 MLP

The performance of the multilayer perceptron (MLP) using single-channel entropy features is reported in Table 3 for both evaluation protocols.

In the subject-dependent setting, the MLP achieves a high average accuracy of 81.41% ± 10.66%, which is slightly lower than the best classical ML approach. The optimal channel varies across subjects, with Pz, T8 and T7 most frequently selected, reflecting the same channel trends observed in the reference paper.

In the cross-subject setting, the MLP reaches an average accuracy of 60.15% ± 6.20%, which is comparable to, but also not higher than, classical ML performance. Pz and AF4 emerge more consistently as the best-performing channels, while the dominance of T8 observed in the subject-dependent case drops.

Overall, the comparison indicates that a shallow MLP does not provide a clear advantage over classical machine learning classifiers when operating on single-channel entropy features. Both approaches exhibit strong subject-dependent performance but struggle to generalize across subjects, and both highlight Pz as the most reliable channel for subject-independent lie detection.

Subject	**LOOCV**		**LOSOCV**	
	Acc. (%)	**Channel**	**Acc. (%)**	**Channel**
S01	74	T7	62	AF4
S02	96	AF4	56	Pz
S03	60	Pz	64	T7
S04	88	T7	60	AF3
S05	74	T7	62	AF3
S06	68	T8	72	T8
S07	82	T7	66	AF4
S08	74	AF3	54	Pz
S09	100	Pz	48	AF4
S10	76	Pz	68	AF4
S11	80	AF4	62	AF3
S12	96	Pz	64	AF3
S13	96	T8	60	Pz
S14	74	T7	62	T8
S15	84	T8	52	T7
S16	80	AF4	60	AF4
S17	70	Pz	48	AF3
S18	90	T8	56	Pz
S19	66	AF4	58	AF4
S20	80	Pz	72	Pz
S21	86	Pz	58	T7
S22	92	Pz	66	Pz
S23	68	AF4	64	T8
S24	72	T7	50	T7
S25	96	Pz	64	Pz
S26	88	T8	58	AF4
S27	88	T8	58	AF3
Mean	81.41	–	60.15	–
SD	10.66	–	6.20	–

Table 3: MLP Performance with best single-channel selection under LOOCV and LOSOCV

4.3 Sequential Model

The performance of the sequential model using single-channel entropy features is summarized in Table 4 for both evaluation protocols.

In the subject-dependent setting, the model reaches an average accuracy of 62.22% ± 13.40%, which is substantially lower than both the MLP and ML classifiers. Accuracies range from below 40% to over 90%, indicating strong inter-subject variability. The optimal channel differs considerably between subjects, with T7 and AF3 most frequently selected, followed by T8 and AF4, while Pz appears only sporadically.

In the cross-subject setting, the average accuracy decreases to 59.93% ± 5.67%, which is comparable to the LOSOCV performance of previous illustrated methods, but still far from robust generalization. Here, T8 becomes the most optimal channel, followed by AF3 and AF4, while T7 and Pz remain rare.

Overall, our sequential model does not demonstrate a clear performance advantage over either classical machine learning classifiers or the MLP.

Subject	LOOCV		LOSOCV	
	Acc. (%)	Channel	Acc. (%)	Channel
S01	46	T8	64	AF4
S02	70	AF4	54	T8
S03	54	AF4	60	T8
S04	62	T7	62	T8
S05	56	T7	64	T8
S06	58	Pz	60	AF4
S07	72	AF3	72	AF3
S08	60	AF3	58	AF4
S09	96	T7	54	T7
S10	62	T7	56	AF4
S11	44	AF3	56	Pz
S12	92	T7	64	AF3
S13	74	T7	68	T8
S14	66	AF3	50	AF3
S15	72	T8	62	T8
S16	38	Pz	58	AF4
S17	58	AF4	58	T7
S18	54	AF3	64	AF4
S19	62	T7	54	AF3
S20	60	AF4	56	AF3
S21	64	AF4	66	AF3
S22	42	T7	54	T8
S23	60	T8	56	T8
S24	76	AF3	64	AF3
S25	58	T8	64	Pz
S26	76	T8	50	AF4
S27	48	AF3	70	T7
Mean	62.22	–	59.93	–
SD	13.40	–	5.67	–

Table 4: Sequential model performance under LOOCV and LOSOCV. For each subject, the best single-channel accuracy is reported.

4.4 Single channel vs channels combination

A consolidated comparison between the best single-channel and best multi-channel configurations across all model families is reported in Table 5 for the subject-dependent setting and in Table 6 for the cross-subject setting. In both evaluation protocols, channel combinations generally lead to substantial and consistent performance improvements over single-channel models.

Under subject-dependent evaluation, all machine learning classifiers exhibit large gains when moving from single-channel to multi-channel configurations, with accuracy improvements ranging from approximately +18% to +22%. The MLP follows the same trend, achieving a +12.7% improvement when using a multi-channel combination. These results indicate that, when using entropy features, multi-channel integration enables the models to better exploit complementary information across electrodes.

A similar effect is observed in the cross-subject setting. Across all machine learning classifiers, multi-channel configurations improve accuracy by up to 27%, while the MLP gains +22.1%, confirming that channel fusion substantially enhances robustness to inter-subject variability.

Notably, the channels that repeatedly emerge as the most informative in the single-channel setting, particularly T7 and T8, are not necessarily those that yield the strongest performance in multi-channel configurations. Instead, other electrodes, such as AF3 and AF4, frequently appear in the best-performing channel combinations, suggesting that channels with weaker individual discriminative power can provide complementary information when combined with others. Among all electrodes, Pz stands out as a consistently informative sensor, appearing in both strong single-channel results and in high-performing multi-channel configurations across all evaluation protocols.

In contrast, the sequential LSTM model, which operates directly on raw EEG signals rather than entropy features, shows only marginal improvements when moving from single-channel to multi-channel configurations, with negligible gains of +2.7% under LOOCV and +2.9% under LOSOCV. This limited benefit suggests that, without explicit feature extraction, raw EEG signals may be too noisy or heterogeneous for effective combined usage, further highlighting the importance of entropy-based representations for multi-channel EEG modeling.

Model	Best single-channel	Acc. (%)	Best multi-channel	Acc. (%)	Δ
SVM-Linear	T8	74.7	AF3+Pz	95.2	+20.5
SVM-RBF	Pz	74.2	T8+AF4	93.9	+19.7
kNN-3	Pz	71.2	AF3+Pz	93.3	+22.1
kNN-5	T8	71.5	AF3+Pz	93.0	+21.5
Naive Bayes	T8	71.9	Pz+AF4	93.0	+21.1
LDA	T8	73.6	AF3+Pz	91.9	+18.3
MLP	Pz	71.1	AF3+Pz	94.1	+12.7
Sequential	T8	52.6	AF3+T7+T8	55.3	+2.7

Table 5: LOOCV average performance, across subjects, comparison between best single-channel and best multi-channel configurations

Model	Best single-channel	Acc. (%)	Best multi-channel	Acc. (%)	Δ
SVM-Linear	Pz	54.6	T7+AF4	75.3	+26.6
SVM-RBF	Pz	53.8	Pz+AF4	75.7	+24.4
kNN-3	AF3	52.2	Pz+AF4	73.1	+20.9
kNN-5	AF3	51.0	AF3+Pz	72.1	+21.1
NB	Pz	52.7	Pz+AF4	75.6	+24.0
LDA	Pz	53.0	AF3+Pz	75.2	+26.0
MLP	AF4	51.6	AF3+T8	73.6	+22.1
Sequential	AF3	52	AF3+AF4	54.9	+2.9

Table 6: LOSOCV performance comparison between best single-channel and best multi-channel configurations

5 Conclusion & future work

The primary objective of this work, reproducing the results of the reference study by Li et al., was successfully achieved. Across both subject-dependent and cross-subject evaluation protocols, our results closely match those reported in the reference paper, thereby confirming the effectiveness and reproducibility of the proposed multi-scale entropy based feature representation.

Beyond reproduction, we extended the original work by investigating deep learning approaches, including a shallow multilayer perceptron trained on entropy-based features and a sequential model operating directly on raw EEG signals. The MLP performed slightly worse than standard machine learning algorithms, while the sequential model yielded substantially poorer results. These findings do not indicate that deep learning is inherently unsuitable for EEG-based lie detection; rather, they highlight the limitations imposed by the small size of the LieWaves dataset and the high noise level of raw EEG signals. The MLP was intentionally designed as a simple architecture. The poor performance of the sequential model suggests that more extensive preprocessing, larger datasets, and more computationally expensive training and hyperparameter optimization are likely necessary for deep learning methods to be effective in this context.

Second, and most importantly, we conducted a comprehensive analysis of multi-channel EEG configurations by evaluating all possible channel combinations. This analysis clearly demonstrates that channel fusion plays a crucial role in improving classification performance. Across all classical machine learning models and the MLP, multi-channel configurations yielded substantial accuracy gains in both subject-dependent and cross-subject settings. Interestingly, the most effective channel combinations were not simply formed by aggregating the best-performing single channels. Instead, channels that were relatively weak when used in isolation—such as AF3 and AF4—were often present in the best-performing multi-channel configurations. This finding highlights the complementary nature of EEG information and suggests that interactions between channels can encode discriminative patterns that are not observable at the single-channel level.

Overall, the results indicate that entropy-based representations combined with multi-channel EEG fusion constitute a robust and effective approach for EEG-based lie detection. At the same time, the strong inter-subject variability observed across all models confirms that generalization to unseen subjects remains a significant challenge.

Several directions for future work emerge from this study:

- **Larger dataset**: the collection of larger and more diverse datasets would be crucial to enable effective training of large deep learning models and to improve generalization across subjects and recording conditions. In parallel, more advanced preprocessing pipelines, such as artifact removal, could reduce noise and enhance discriminative neural patterns.
- **Multi-modal fusion**: the combination of EEG with other physiological signals (ECG, GSR, eye tracking) could improve detection robustness, fusing features at early (raw signals) or late (decision) stages.

- **Model architectures**: from a modeling perspective, future studies could explore architectures specifically tailored to EEG data, including attention-based temporal models, graph neural networks that explicitly encode electrode topology, and hybrid approaches combining handcrafted features with learned representations.

In conclusion, this work confirms the effectiveness of entropy-based EEG features for lie detection designed by Li *et al.* and provides strong evidence that multi-channel integration is a key factor for improving performance and robustness in EEG-based deception detection systems.

References

[1] B. N. Taha, M. Baykara, and T. B. Alakuş, "Neurophysiological approaches to lie detection: A systematic review," *Brain Sciences*, vol. 15, no. 5, 2025, ISSN: 2076-3425. DOI: `10.3390/brainsci15050519`. [Online]. Available: `https://www.mdpi.com/2076-3425/15/5/519`.

[2] J. Li et al., "A resource-efficient multi-entropy fusion method and its application for eeg-based emotion recognition," *Entropy*, vol. 27, no. 1, 2025, ISSN: 1099-4300. DOI: `10.3390/e27010096`. [Online]. Available: `https://www.mdpi.com/1099-4300/27/1/96`.

[3] J. Li et al., "A novel multi-scale entropy approach for eeg-based lie detection with channel selection," *Entropy*, 2025.

[4] M. Aslan, M. Baykara, and T. B. Alakuş, *Liewaves dataset: Eeg signals for deception detection*, Dataset description provided by the authors, 2024.