INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 2, Month: April, Year: 2025
Article Number: 7008, https://doi.org/10.15837/ijccc.2025.2.7008

A multilayer perceptron neural network prediction approach to polygraph scoring

Dana Rad, Csaba Kiss, Gavril Rad, Nicolae Paraschiv, Marius Bălaş

Dana Rad
1. Petroleum-Gas University of Ploiești, Romania
2. Aurel Vlaicu University of Arad
Faculty of Educational Sciences Psychology and Social Work
Center of Research Development and Innovation in Psychology, Arad, Romania
dana@xhouse.ro

Csaba Kiss
Hyperion University of Bucharest, Romania
kiss.csaba@expertpoligraf.ro

Gavril Rad
Aurel Vlaicu University of Arad
Faculty of Educational Sciences Psychology and Social Work
Center of Research Development and Innovation in Psychology, Arad, Romania

Nicolae Paraschiv
Petroleum-Gas University of Ploiești, Romania
nparaschiv@upg-ploiesti.ro

Marius Bălaş
Aurel Vlaicu University of Arad
Faculty of Engineering, Arad, Romania

Abstract

Years of studies have consistently demonstrated that people's capacity to detect deceit is no better than chance. For law enforcement officers, accurate deception detection is critical. The traditional polygraph examination is now the sole standardized and reliable method for detecting deceit. There are several standardized scoring protocols (Lafayette Polygraph System 11.8.6) to Control Question Technique (CQT) Polygraph examinations: PolyScore, OSS-2, OSS-3 and manually scoring. Due to the ongoing controversy over which scoring system performs better in terms of avoiding false positive and false negative errors, this study introduces a Multilayer Perceptron Neural Network (MLP) prediction approach to Polygraph deception scoring utilizing manually scored examination data. A MLP was trained to predict high and low deception scores in 400 offender data, based on the most predictive psychophysiological indicators found in the scientific literature: amplitude of electrodermal reaction (ARED), amplitude of blood pressure in brachial pulse (ATAB), change of base line level in chest breathing (MNBRT) and difference of altitude between breathing cycles (DIFA). The model predicted the deception level of the 400 offenders with a correct classification rate (CCR) of 80%, result consistent with the prediction accuracy reported in the recent literature. The MLP neural network modeling results showed that based on the four psychophysiological indicators ARED, ATAB, MNBRT and DIFA there is an 80% correct classification rate of high and low deception scores received by insincere subjects. The key outcome of this study suggests that MLP represents a robust approach to identify deception in manually scored polygraph examinations.

Keywords: Polygraph, scoring system, Multilayer Perceptron Neural Network (MLP), deception detection.

1 Introduction

This paper presents a novel approach using a Multilayer Perceptron (MLP) Neural Network for the prediction of deception scores in polygraph examinations, in line with the journal's focus on Artificial Intelligence, specifically deep learning, and the integration of advanced computational methods in various applications. In many fields, deception detection accuracy and dependability are crucial, especially in law enforcement, where conventional polygraph techniques are still widely used despite continuous discussions over their efficacy. According to the journal's focus on deep learning, this use of MLP illustrates how artificial intelligence might improve the predicted accuracy of intricate psychophysiological tests.

Artificial intelligence has been shown to be effective in maximizing predictive accuracy and improving the dependability of computational systems in a variety of domains, including real-time intelligence systems and QoS prediction, according to recent studies by Wahsheh et al. (2021) and Albu et al. (2014). In a similar vein, the current study uses an MLP neural network to train the model on important psychophysiological signs in order to solve the difficulties associated with evaluating polygraph deception. The examination data is manually scored. According to Prepeau et al. (2017), this strategy will contribute to the advancement of cutting-edge AI applications, especially deep learning, which will boost the efficiency of computational systems in real-world situations (Sinescu et al., 2009; Toader et al., 2023).

It's a popular belief in today's society that one can tell someone is honest simply by looking at them. Physical cues are not always dependable markers of dishonesty, despite the fact that they might provide some insight into someone's emotional condition. As a result, novel technologies—like artificial intelligence-based lie detectors—have been developed with the intention of more accurately identifying dishonesty.

Artificial intelligence (AI)-based lie detectors are more useful in large-scale scenarios like airport security checks and employment candidate assessments because of their many benefits, including speed and ease of administration. Additionally, these systems can be configured to pick up on minute physiological alterations that the human eye or conventional polygraph machines could miss. Notwithstanding these advantages, questions concerning the precision and dependability of AI-based lie detectors still exist. Although supporters contend that they are more accurate than conventional polygraph examinations, experts are still debating the veracity of these assertions. While some academics are concerned about the possibility of prejudice and discrimination in their use, others wonder if these systems are indeed successful at detecting deceit.

However, the field of deceit detection has advanced significantly with the creation of AI-based lie detectors. Future technological advancements are certain, hence it is imperative that scholars and decision-makers carefully consider the ethical and factual ramifications of these cutting-edge inventions. Because AI-based lie detectors can be utilized in situations where traditional polygraph examinations are impractical, they have become more and more popular in recent years. Because of the technology's apparent modernity and complexity, these systems are being utilized more frequently, for instance, in private job interviews, border crossings, loan checks, and insurance fraud claims, among other contexts (Wang, Shi & Liu, 2021).

These technologies are being widely adopted by businesses and governments to evaluate the reliability of customers, employees, locals, immigrants, and international guests. Applications for visas, employment screening, security clearance evaluations, and other situations have all made use of these systems (Giansiracusa, 2021; Sánchez-Monedero & Dencik, 2022; Oravec, 2022; Brouwer, 2021; Elkins, Gupte & Cameron, 2019). Customization to meet the requirements of various applications is one benefit of AI-based lie detectors. For instance, some systems might be made to examine spoken answers to inquiries, while others might be more concerned with examining physiological reactions like skin conductance or pulse rate. Furthermore, such algorithms might be designed to identify particular linguistic or behavioral patterns linked to dishonesty.

These technologies have advantages, but there are also questions regarding their validity and dependability. Opponents of AI-based lie detectors contend that its accuracy might not be high enough to justify their application in critical decision-making scenarios, like immigration and employment screening. Concerns about privacy and the security of personal data are present, in addition to possible bias in the creation and application of these technologies. Polygraph techniques are employed not just in criminal investigations but also extrajudicially to assess an individual's integrity and dependability. One well-known example is the use of polygraph tests in some companies' hiring procedures for new hires.

One of the first studies to investigate the use of polygraph testing in pre-employment screening was Ash (1971), and his results implied that the test might reliably identify candidates who were likely to act dishonestly. In a similar vein, Nagle (1983) discovered that polygraph examinations were an effective means of vetting candidates for high-security jobs. Cunningham (1989) looked into the aviation industry's use of polygraph testing and discovered that it may be useful in identifying people who were a security concern. Zafran and Stickle (1984) also documented the effective application of polygraph examinations in federal law enforcement agencies' hiring procedures. More recently, Nevins (2004) investigated the application of polygraph testing in the financial sector and discovered that it would be a useful method for detecting people who were a risk to the company.

From the first concepts of polygraph testing, modern deception detection systems have advanced to include more complex data analysis and machine learning algorithms. However, experts and practitioners have disagreed over how accurate these technologies are. Research has indicated that conventional methods of detecting dishonesty, like the original polygraph concept and the polygraph itself, are not entirely accurate. Slowik (2013) and Matte (1996) discovered that the polygraph's accuracy was barely superior to random chance. Raskin (1987) and Cross and Saxe (1993) both noted low polygraph accuracy rates. Matte (2007) and Horvath (2020) discovered that the subjective interpretations of polygraph charts by examiners were not trustworthy. Furthermore, Slowik (2020) and Lapadusi and Dobreanu (2014) discovered that the polygraph's accuracy dropped when applied to real-world situations.

Gordon (2016) investigated whether computational text analysis could identify dishonesty at accuracy rates that were on par with human experts. In a meta-analysis of forty research on the polygraph's accuracy, Raskin and Kircher (2014) discovered that the test's ability to identify deceit was just marginally superior to chance. Nelson (2015) also examined the research on the polygraph and came to the conclusion that its validity and accuracy were constrained. More recently, Raskin et al. (2019) discovered that a machine learning algorithm performed better than a standard polygraph test in their study comparing the accuracy of the two methods. The possibility of machine learning algorithms for deception detection was also investigated by Widacki (2020), who discovered that these systems could identify dishonesty with great accuracy.

Machine learning algorithms have the potential to increase accuracy, according to recent study, but further studies are required to completely assess their efficacy in practical contexts. Future research on deceit detection appears to be headed in a promising route with the application of machine learning algorithms and extensive data analysis. Accuracy improvements appear possible thanks to recent advancements in machine learning algorithms and complicated data processing. In controlled laboratory settings, machine learning systems were found to be able to accurately identify dishonesty by Gordon et al. (2006) and Bhutta et al. (2015). In addition, O'Shea et al. (2018) found that a machine learning method produced high accuracy rates when used in real-world situations. Furthermore, Oswald (2020) and Raskin, Kircher, and Honts (2019) discovered that machine learning algorithms could raise the precision of conventional polygraph testing.

In conclusion, current advancements in complicated data analysis and machine learning algorithms have showed promise in enhancing accuracy, even while old deception detection tools have proven to be unreliable. Nevertheless, more investigation is required to properly assess these new technologies' efficacy in practical contexts.

2 Materials and Methods

2.1 Participants

The 400 participants were chosen at random from a group of 1072 dishonest individuals who had committed several offenses and have been evaluated with Polygraph test by expert polygraph examinators from 10 polygraph laboratories of the Romanian Police, coordinated by Dr. Csaba Kiss. All 400 participants are criminals who have committed serious crimes on a recidivist basis and all confessed their crimes and gave their consent that their aggregated data to be used in scientific research. The information that was used in the present research was gathered from a minimum of three diagrams related to each test performed during the Polygraph assessment. The investigated individuals were 90% men and 10% women, with an age span between 18 and 65 years old and an average age of 32 years. The mean length of finalized educational studies was and 8.6 years, a rather low level of education.

2.2 Measures

The method employed in this study is forensic psychophysiology examination, commonly known as polygraph testing, conducted over a three-year period from 2004 to 2007. These examinations were carried out in 10 polygraph laboratories within the Romanian Police, staffed by polygraph examiners with nearly equal levels of seniority (one staff member joined in 1998, eight in 1999, and one in 2000). The polygraph method is non-intrusive and adheres to the principles of personal integrity and the presumption of innocence.

Two families of polygraph techniques were utilized: the Modified General Question Technique (MGQT) and the Air Force Modified General Question Test (MGQT-AIR FORCE). These techniques are derived from various modifications of the General Question Technique and the Zone Comparison Technique (Backster, 1963). Both MGQT and MGQT-AIR FORCE tests were administered following all the stages of a polygraph examination, including numerical scoring, a 7-step scale, and OSS scale, all evaluated manually.

Physiological arousal factors, including heart rate, blood pressure, respiration, sweat, and skin conductivity, are measured by MGQT and MGQT-AIR FORCE. The fundamental idea behind polygraph testing is that when a subject is telling the truth as opposed to lying, these physiological reactions will change.

The list of the 22 parameters and their descriptive statistics are presented in table 1.

The ground-truth data for this study was established based on offenders' confessions obtained during the posttest interviews following all polygraph examinations. These confessions provide a reliable baseline for validating the results of the neural network model.

2.3 Procedures

After reviewing the scientific literature, we concluded that the level of deception score in manual Polygraph examinations scoring systems and not only, is highly predicted by the following indicators: amplitude of electrodermal reaction (ARED), amplitude of blood pressure in brachial pulse (ATAB), change of base line level in chest breathing (MNBRT) and difference of altitude between breathing cycles (DIFA) (Cook & Mitschow, 2019). A strong theoretical basis and empirical support are provided for our investigation by the general applicability and standardization of manual expert scoring (Gogate, Adeel & Hussain, 2017; Pérez-Rosas et al., 2015).

The novelty of our approach is that there has never been reported a study that employs NN algorithms to estimate the deception scores in polygraph testing.

	N	Minimum	Maximum	Mean	Std. Deviation
ARED	400	.00	20.80	6.4430	4.14712
ATAB	400	.00	14.70	2.9280	1.98674
ATAD	400	.00	18.00	.8063	1.68062
ART	400	.00	6.40	1.7180	.89928
ARA	400	.00	6.60	1.8945	.97460
MNBRA	400	.00	6.20	.2845	.55274
MNBRT	400	.00	8.10	.2902	.63592
IR	400	.00	1.54	.5672	.25061
LLRT	400	128.00	1184.00	379.3050	148.84446
LLRA	400	176.00	1480.00	442.2675	205.29060
LRED	400	76.00	1440.00	329.0175	205.23037
TRED	400	.00	87.00	13.7027	5.48503
TTAB	400	.00	37.00	11.9035	5.36883
TTAD	400	.00	24.60	4.2175	4.32322
RR	400	.00	32.00	18.5505	4.49283
RC	400	36.00	152.00	84.7300	16.39020
TSTOPR	400	.00	20.00	1.2515	3.19617
TSTOPRA	400	.00	20.00	1.2295	3.13461
REV	400	.00	1.00	.0975	.29701
EDA	400	63.00	886.00	169.6400	89.67393
DIFA	400	.00	19.00	.9867	1.47295
TDIFA	400	.00	27.30	8.0985	6.27655
Valid N (listwise)	400				

Table 1: Descriptive statistics of the 22 physiological parameters

The current research uses a Multilayer Perceptron (MLP) modeling technique to predict the level of deception scores (high deception and low deception) of offenders obtained by manual scoring diagrams, based on recognized parameters the amplitude of the electrodermal reaction (ARED), amplitude of blood pressure in brachial pulse (ATAB), change of base line level in chest breathing (MNBRT), and difference of altitude between breathing cycles (DIFA). This research was motivated by the demand for technology integration into psychological examinations like Polygraph examinations that are using psychophysiological measurements.

The choosing of these particular 4 physiological parameters: amplitude of electrodermal reaction (ARED), amplitude of blood pressure in brachial pulse (ATAB), change of base line level in chest breathing (MNBRT) and difference of altitude between breathing cycles (DIFA) was decided upon the results of a multiple linear regression analysis of all physiological parameters included in this investigation: DIFA, the duration of the electrodermal reaction (TRED), ARED, abdominal breath line length (LLRA), arterial tension amplitude of the distal pulse (ATAD), heart rhythm (RC), voluntary repeated acts (REV), duration of brachial pulse arterial tension (TTAB), changing of the basic level of abdominal breathing (MNBRA), the ratio of inspiration to expiration (I/E), the average value of the electrodermal reaction (EDA), thoracic breath line length (LLRT), reactive patterns (PATTR), duration of distal pulse arterial tension (TTAD), ATAB, respiratory rhythm (RR), erratic breathing (RE), abdominal respiratory stop (TSTOPRA), average amplitude of abdominal breathing (ARA), the length of the electrodermal reaction (LRED), MNBRT, and thoracic respiratory stop (TSTOPR).

All 22 physiological parameters accounted for 28% variance in the brut expert score which ranged from -1 to -2, as in the standard manual scoring procedure, with an $F = 8.055$ at $p < 0.01$. The standardized $Beta$ coefficients that obtained significance level of $p < 0.01$ were: ARED ($Beta = -.203$, $p < 0.01$), ATAB ($Beta = -.256$, $p < 0.01$), MNBRT ($Beta = -.135$, $p < 0.05$), and DIFA ($Beta = -.168$, $p < 0.05$). An additional parameter PATTR has obtained significant predictive level ($Beta = -.186$, $p < 0.01$), but it was discarded from the neural network predictive analysis due to the fact that reactive patterns were registered with only 3 values: 0, 1 and 2, being a coded parameter that was not directly measured. The Beta coefficients have negative loadings due to the fact that the expert score is negatively registered, being a deception score, according to the standardized scoring procedure. The expert score was further processed into 2 distinct categories, 1 for low deception scores including brut expert scores from -1 to -5 (frequencies representing a cumulative 52% of total data) and high deception scores ranging from -6 to -9 (frequencies representing a cumulative 48% of total data).

Predictor variables are continuous variables with the following descriptive statistics: for ARED scores ranged from 0 to 20.80, with a mean of 6.44 and a standard deviation of 4.14; for ATAB, scores ranged from 0 to 14.70, with a mean of 2.92 and a standard deviation of 1.98; for MNBRT, scores ranged from 0 to 8.10, with a mean of 0.29 and a standard deviation of 0.63, and for DIFA, scores ranged from 0 to 19.00, with a mean of 0.98 and a standard deviation of 1.47. The dependent variable, or the output variable in the neural network algorithm is represented by manual expert scores that was coded with 1 for low deception and 2 for high deception, all subjects in this research are insincere subjects, with offending background.

Due to the continuing advancement of artificial intelligence and advanced algorithms, data-driven methodologies have become more and more adopted as a successful modeling tool in recent years (Popescu et al., 2009; Rad et al., 2022). The Multilayer Perceptron MLP is often utilized due to its straightforward construction and exceptional capability for function approximation (Marouf et al., 2019). One or more layers of neurons make up a multilayer perceptron, a type of neural network. Predictions are created on the output layer, often referred to as the visible layer, after data is received from the input layer and abstracted to varying degrees by one or more hidden layers. MLPs are excellent for classification prediction problems using labeled inputs. They are also suitable for regression prediction problems where a real-valued quantity must be predicted given a collection of inputs.

A particular kind of feedforward artificial neural network is called a multilayer perceptron (MLP). The MLP consists of a signal-receiving input layer, a decision-making output layer, and an arbitrary number of hidden layers that act as the MLP's true computational engine. In the perceptron, learning occurs by changing connection weights based on the degree of error in the output relative to the expected result after each piece of input is processed. In this example of supervised learning, backpropagation, an extension of the least mean squares method, is employed. Any continuous function may be approximated by MLPs with a single hidden layer. Frequently, multilayer perceptrons are employed to address supervised learning problems. They learn to model the correlations or dependencies between those inputs and outcomes by practicing on a variety of input-output combinations. In order to improve accuracy, training requires changing the model's parameters, or weights and biases. The weight and bias adjustments are made via backpropagation in proportion to the error, which may be measured in a number of different ways, including root mean squared error (RMSE).

We used an MLP neural network strategy to forecast the amount of deception scores (high deception and low deception) of 400 offenders, in terms of proper classification, as opposed to the traditional multiple linear regression statistical processing of data. The following research topic is the specific subject of this study: Using only the four psychophysiological measurements ARED, ATAB, MNBRT and DIFA, is the MLP neuronal network modeling approach a reliable prediction technique for deception scores?

The measurements were standardized in the sense that the adjustment buttons were positioned at the beginning of the measurements next to the same values for each route and subject evaluated. Examinations and evaluations were performed after each device was calibrated with a calibrator and the screen of each computer was also calibrated to avoid any distortion generated by the device.

Apparatus and software LX 4000 type devices were used, each equipped with 2 pneumograph tubes, a galvanometer, a plethysmograph and a blood pressure sleeve. Also, the software used was LX 9.9.5 which incorporates the Calipers section useful for accurate measurements. The most intense physiological answers recorded to a single relevant question of each examination were selected. The comparison questions associated with them were also used to establish the expert scores (by numerical scoring). The measurements were performed in compliance with the rules established by the APA (American Polygraph Association) at intervals of at least 25 seconds. Following the completion of the data extraction, we built a database in SPSS V.26 and utilized its MLP feature to predict criminal offenders' high and low deception scores using NNs modeling.

3 Results

MLPs are the most fundamental deep neural network models, consisting of a succession of fully linked layers. We have further employed MLP machine learning to overcome the requirement of high computing power required by modern deep learning architectures for the prediction of the level of deception in criminal offenders that confessed their crimes, based on only 4 parameters ARED, ATAB, MNBRT and DIFA out of 22 parameters described in the procedures subsection.

In order to employ the MLP technique, we have used the path Analyze/Neural Networks/Multilayer Perceptron in SPSS V.26 program. In this study, we used a typical multiple-input, single-output MLP. We have proceeded with designing the architecture as follows. First, we have selected the dependent variable as expert score, then the 4 parameters: ARED, ATAB, MNBRT and DIFA were selected in the covariates section. We have chosen the standardization method for rescaling of covariates. In the partition section, we have chosen to randomly assign cases based on relative number of cases: 70% training and 30% testing, with no partitioning variable to assign cases. Thus, out of a total of 400 instances, 281 cases (or 70.3 percent) were assigned to the training sample, while 119 cases (or 29.7 percent) were assigned to the testing sample. ARED, ATAB, MNBRT and DIFA are four independent variables that are covariates in the network information output.

The goal of designing our MLP neural network was to maximize performance, minimize computational resources during training, maximize the level of automaticity by minimizing the number of decisions that need to be made by a human during the design process, and to minimize the model's complexity, specifically the network's size. As a result, we have chosen automatic architecture selection. As a result, the automated architecture design option suggested using the tangent hyperbolic (tanh) function for hidden layer activation and the Softmax function for output layer activation. The only non-linear activation functions for the hidden layer in SPSS v.26 are tangent hyperbolic (tanh) and sigmoid. The tanh (tangent hyperbolic) function's output always varies between -1 and +1, and it, like the sigmoid function, displays an s-shaped graph, indicating a non-linear function. One advantage of utilizing the tanh function over the sigmoid function is that it is zero centered, which makes optimization much easier. SPSS V.26 has four alternative types of output layer default activation functions: identity, softmax, tanh, and sigmoid. The input layer only stores the input data and does not perform any calculations. As a result, no activation function was employed. To forecast a multinomial probability distribution, the softmax function was employed as the activation function in the MLP model's output layer.

Figure 1 shows the MLP architecture with all the layers mentioned previously.

In the model summary, we discovered a cross entropy error of 127.437 and a training sample percentage of inaccurate predictions of 21%. A cross entropy error of 55.891 and a percentage of inaccurate predictions of 20.2% were found in the testing sample. Thus, 80% of the categorized data are appropriately allocated to the dependent variable's expert score value, which is a considerable rate of accurate predictions in the testing sample. A comparison of the suggested MLP neural network-based approach's performance to the 82 percent prediction accuracy reported in previous literature has yielded similar results (Gogate et al., 2017). Additionally, we show the correct classification percentage for each data point in the training and testing samples in Table 2. With an overall accurate percent of 79 percent, data point 2 (high deception) had the lowest correct percent in the training sample, at 46.8 percent. With an overall accurate percent of 79.8 percent, data point 2 (high deception) in the testing sample had the lowest correct percent, at 51.4 percent. As evidence of the effectiveness of our MLP classification model, the anticipated pseudo-probability, sensitivity-specificity, cumulative gain, and lift chart outputs are shown in Figure **??**.

We also discuss the performance of our model in terms of the gain/lift analysis, which helps to select the best predictive model among several competing models. To evaluate the effectiveness of a classification model, gain and lift charts are utilized. They evaluate how much better a person could expect to do in the absence of a prediction model. The slope shown in Figure 2 gradually decreases as the number of records that truly belong to the class of interest to add decreases and the model's

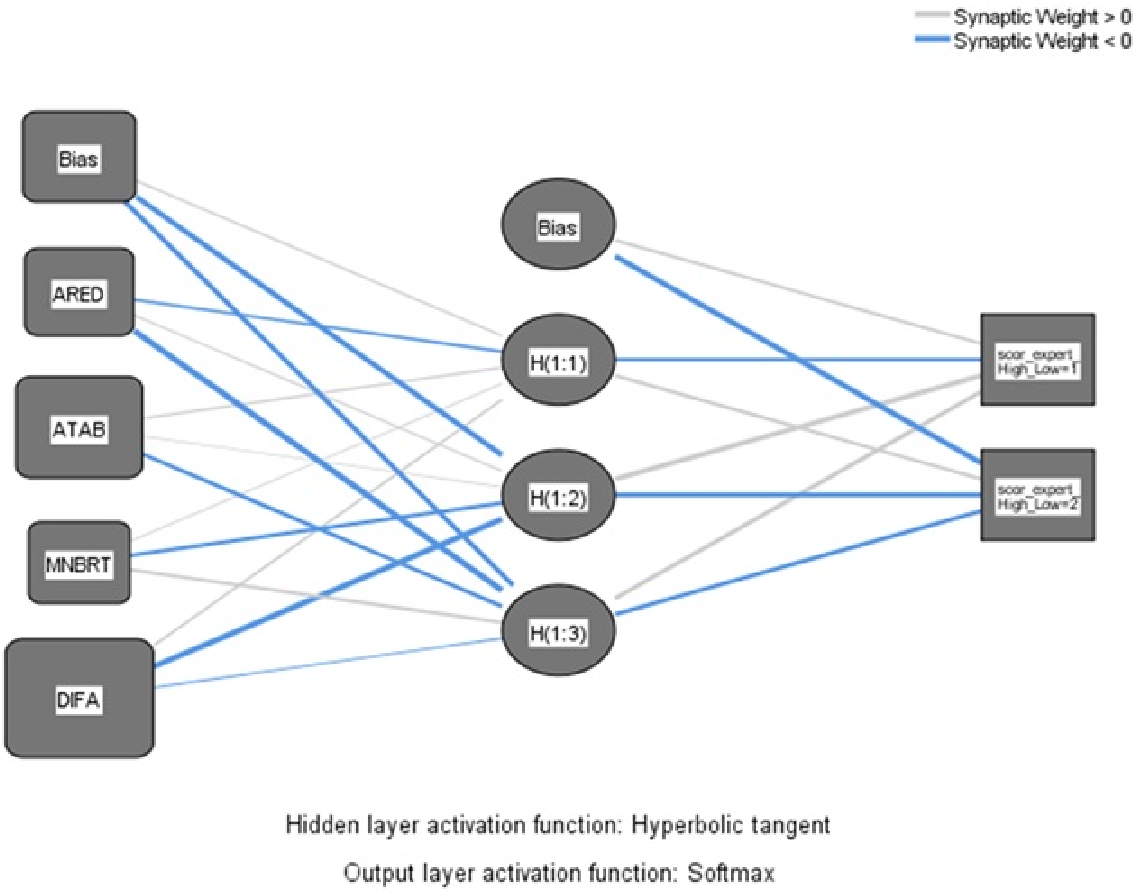

Figure 1: The MLP neural network architecture

Sample	**Observed**	**1**	**2**	**C** **Percent Correct**
Training	1 *(expert score – low deception)*	185	17	91.6%
	2 *(expert score high deception)*	42	37	46.8%
	Overall Percent	80.8%	19.2%	79.0%
Testing	1 *(expert score – low deception)*	77	7	91.7%
	2 *(expert score high deception)*	17	18	51.4%
	Overall Percent	79.0%	21.0%	**79.8%**

Dependent Variable: expert score (low deception – 1, high deception – 2)

Table 2: Classification

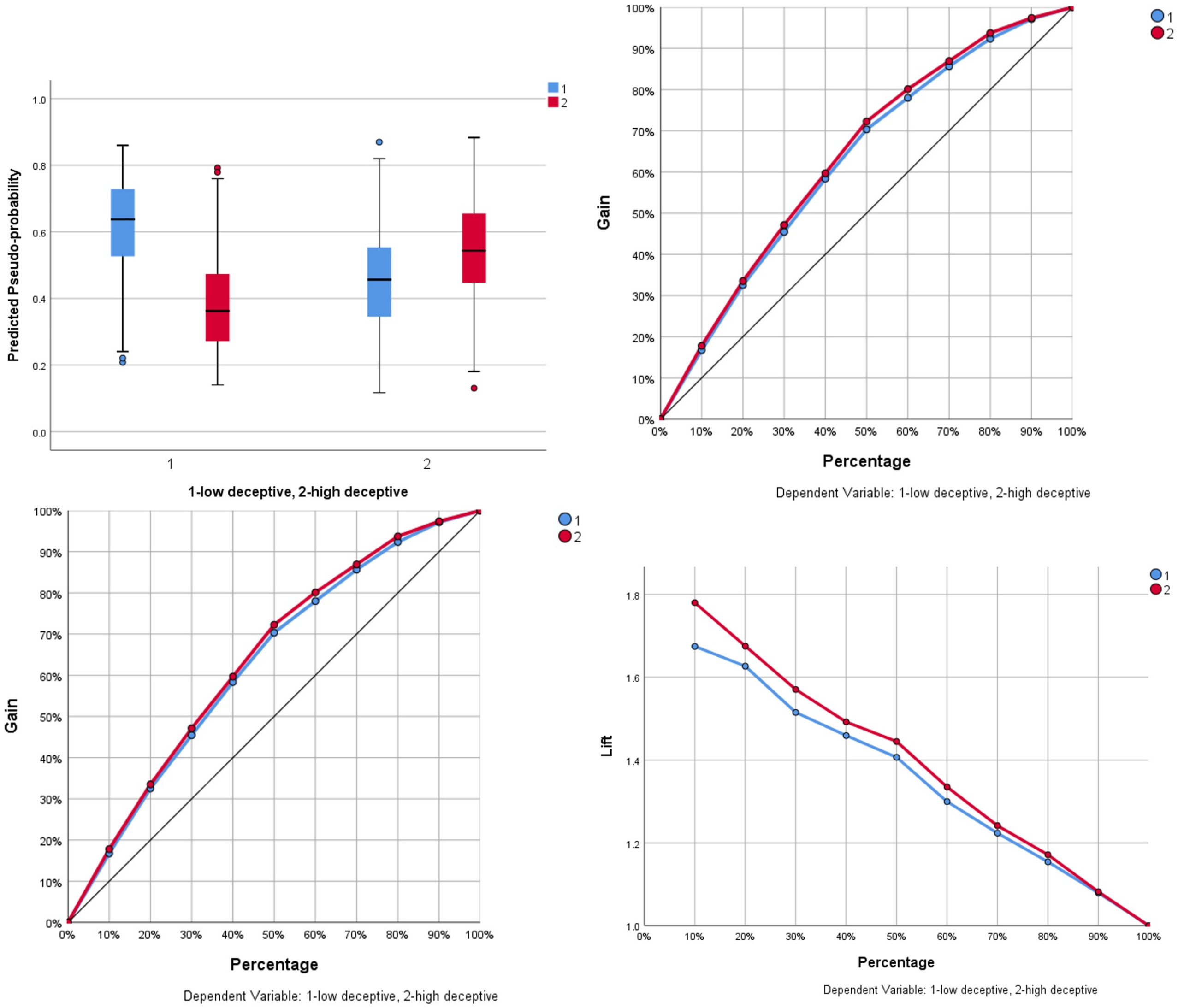

Figure 2: Pseudo-probability, sensitivity-specificity, cumulative gain, and lift chart outputs for Dependent Variable: expert score (low deception – 1, high deception – 2)

ability to provide an advantage decrease, as indicated by the lift chart. Our MLP works effectively, producing a notable "lift" for a sizable portion of the ranking data. When applying the MLP model, the area under the curve (AUC) value exhibits great overall performance. We received a score of 0.81 for data point 1 (low deception scores) and the same value (0.81) for data point 2 (high deception scores). An AUC of 0.5 often indicates no discrimination, whereas values between 0.7 and 0.8 are regarded as good, between 0.8 and 0.9 as excellent, and values beyond 0.9 as remarkable.

	Importance	**Normalized Importance**
ARED	.185	54.4%
ATAB	.340	100.0%
MNBRT	.140	41.3%
DIFA	.335	98.7%

Table 3: Independent variable importance

The importance score in table 3 and figure 3 indicates the contribution of each variable to the model's overall accuracy in predicting the dependent variable. In this case, the variable with the highest importance score is ATAB (0.340), followed by DIFA (0.335), ARED (0.185), and MNBRT (0.140). The normalized importance score expresses each variable's importance relative to the most important variable (in this case, ATAB), which is assigned a value of 100%. According to this metric,

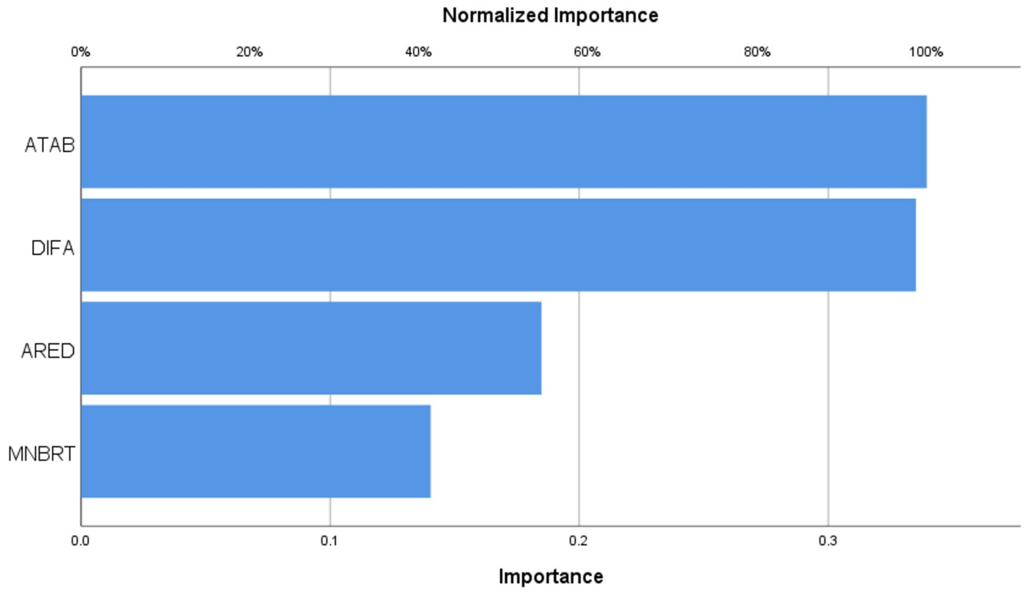

Figure 3: Independent variable importance.

ATAB accounts for 100% of the model's predictive power, while DIFA accounts for 98.7%, ARED for 54.4%, and MNBRT for 41.3%.

Overall, these results suggest that ATAB is the most important variable in predicting the dependent variable, followed closely by DIFA, ARED and MNBRT appear to have relatively less influence on the model's predictive accuracy

	Predicted: Low Deception	**Predicted: High Deception**
Actual: Low Deception	185	17
Actual: High Deception	42	37

Table 4: Confusion matrix

The confusion matrix presented in Table 4 provides a detailed evaluation of the performance of the Multilayer Perceptron (MLP) model by comparing actual deception scores to predicted deception scores. It reveals that the model correctly predicted low deception in 185 cases (true positives) and high deception in 37 cases (true negatives). However, it incorrectly predicted low deception in 42 high deception cases (false positives) and high deception in 17 low deception cases (false negatives). Based on the confusion matrix, the model's overall accuracy is 79%, meaning it correctly predicts deception scores in 79% of cases. The precision, or the proportion of true positive predictions for low deception, is 81.5%, indicating that when the model predicts low deception, it is correct 81.5% of the time. The recall (sensitivity), representing the proportion of actual low deception cases correctly identified, is 91.6%, showing that the model accurately detects 91.6% of all low deception cases. In contrast, the model's specificity, or ability to correctly identify high deception, is 46.8%, meaning it struggles more with detecting high deception cases. The F1 Score, a balance between precision and recall, is 86.2%, indicating that the model performs well in balancing these two aspects. Overall, the model is highly effective at detecting low deception but requires improvement in identifying high deception cases.

4 Discussion

This study involves an early assessment and analysis of polygraph data. We obtained accuracy rates that were comparable to those reported by other algorithms and manual scoring using a reasonably straightforward procedure. The fact that we are able to construct a broad variety of alternative models that take a variety of parameters into consideration while producing identical findings shows how complex the evaluation and/or categorization of examinee dishonesty is.

Providing precise and statistically sound classification systems with low false positive and false negative rates is the aim of automated scoring algorithms for polygraph data. Numerous research published in the polygraph literature starting in the 1970s asserted to show the efficacy of automated categorizing methods and algorithms for assessing polygraph charts. Dollins, Kraphol and Dutton (2000) claim that when inconclusive data are included, the accuracy of five different computer algorithms on dishonest people ranges from 73 percent to 89 percent, and from 91 percent to 98 percent when they are deleted.

Our results demonstrate that 80% of the categorized data are appropriately allocated to the dependent variable's expert score value, which represents a considerable rate of accurate predictions in the testing sample. A comparison of the performance of the proposed MLP neural network-based technique to the 82 percent prediction accuracy reported in earlier research showed similar results (Gogate et al., 2017). The scientific literature on the subject of polygraph scoring technique also supports the conclusion that ATAB is the most significant predictor of the Polygraph score (Pasca, 2012; Saeed et al., 2022).

The main limitation of our research is represented by the scarcity of nondeception examples, which is one likely explanation for the acceptable classification performance.

Another limitation may be the high degree of variability and measurement errors found in real-world polygraph data when there are inadequate standards for data collection and recording. The problems with question inconsistencies, answer variability within and between people, and potential learning effects are shown by our exploratory data analysis. It is not always obvious where variations in answers occur, whether we are dealing with habituation or comparing semantically distinct items throughout the charts. In our method, we have averaged the relevant and control replies and then looked at their difference because our data queries are semantically diverse and no consistent ordering inside and across charts could be built. The methods employed by OSS and PolyScore take a similar approach.

This work has the potential to be reinterpreted and expanded in a variety of ways. More characteristics might be retrieved and investigated. So far, these efforts have not resulted in considerably lower mistakes, raising the issue of how far this strategy may go beyond what has previously been documented. The order of the questions needs to be taken into account. A mixed-effects model with repeated measures, where the repetitions would be measurements over many charts, is an additional choice.

For either single occurrences or security screening, it has not yet been able to conduct a fully independent examination of computer scoring systems on a large enough sample of instances to allow one to safely assess the validity and accuracy of these algorithms. One might contend that neural network algorithms should be better at data analysis since they are capable of tasks that even experienced examiners find difficult, such as filtering, transformation, computing signal derivatives, manipulating signals, and looking at the big picture rather than just adjacent comparisons. Our results bring promising evidence for a future MLP neural network scoring feature integration into the standardized Polygraph scoring systems, especially needed in cases of diagnostic impossibility, thus assisting clinical scoring of polygraph diagrams when there is an impossibility assessment offered by current scoring systems OSS and Polyscore.

However, the effectiveness of both numerical and electronic methods is still significantly dependent on the examination's pre-test phase. The quality of information recorded is inextricably linked to how effectively examiners design the questions (Walczyket et al., 2013). The eventual potential of computerized scoring systems depends on both the consistency of the test formats that the systems are designed to handle and the quality of data that is available for system development and implementation. We believe there is space for significant modifications to be made to the current numerical scoring.

Traditional lie detection methods focus on behavioral or psychophysiological signs of dishonesty, such as the polygraph, voice stress analysis, or specific interrogation techniques The issue of whether it would be able to directly recognize deception in the area of the body where it is generated—the brain—arose with the development of neuroimaging technology. Unfortunately, much of the research in this area was relatively arbitrary and ignored the body of information regarding methodological traps that were hotly debated in the scientific community in relation to the polygraph (Bashore & Rapp, 1993).

Because of this, there are significant differences across neuroimaging research on deception in terms of the experimental paradigm (the interrogation technique), the data analysis techniques, and the methodologies used to make individual diagnoses. Additionally, the majority of research made use of fabricated laboratory conditions that are very unlike from real-world applications. As a result, neuroimaging methods are not currently useful for identifying dishonesty in specific field scenarios. However, new developments like multivariate pattern analysis may soon result in fresh neuroimaging applications that can enhance current methods for spotting dishonesty or hidden information (Gamer, 2014; Lo, Fook-Chong & Tan, 2003).

The investigation of the phenomena of self-deception and lying could benefit from further research using screening tests or standardized laboratory situations. These methods ensure that questions are asked in the same sequence and vary only slightly between individuals and charts. Additionally, future studies should consider the use of electroencephalography (EEG) brain mapping techniques to add a rigorous measurement dimension to the psychophysiological measurements in classical polygraph measurements.

EEG brain mapping refers to a collection of separate techniques for quantified EEG analysis that measure the electrical activity of the brain (Nuwer, 1990a; Nuwer, 1990b). These techniques provide a more objective way to measure physiological responses related to deception and self-deception (Cook & Mitschow, 2019; Happel, 2005). For instance, researchers can analyze changes in brain waves or neural activity when individuals engage in deceptive behavior, which can provide insight into the cognitive processes involved in lying (Arasteh, Moradi & Janghorbani, 2016; Bhuvaneswari & Kumar, 2015).

Several studies have utilized EEG brain mapping techniques to investigate deception and self-deception. Meijer and Verschuere (2017) used EEG to examine brain activity during deception and found that the brain's prefrontal cortex was more active when participants were lying compared to telling the truth. Gamer (2014) also found that lying was associated with increased activity in the prefrontal cortex, particularly in regions involved in cognitive control and decision-making. Other studies have focused on identifying specific EEG markers that can reliably indicate when someone is lying. Kohan, Nasrabadi, and Shamsollahi (2020) developed a new algorithm that analyzed EEG signals to detect deception with high accuracy. Lakshan et al. (2019) used EEG to identify differences in brain activity between deceptive and truthful responses in a mock crime scenario.

In addition to investigating deception, EEG brain mapping techniques can also shed light on the mechanisms underlying self-deception. EEG was utilized by Nortje and Tredoux (2019) to investigate the connection between cognitive dissonance and self-deception. Researchers discovered that those who self-deceived had lower brain reactions to stimuli that caused dissonance, which may indicate that these people are less conscious of the inconsistencies between their beliefs and actions. EEG was utilized by Daneshi Kohan and associates (2020) to examine the connection between various forms of cognitive control and deceit. According to their findings, distinct brain patterns are linked to various forms of dishonesty, and electroencephalography (EEG) holds potential as a means of identifying deceit in forensic situations.

All things considered, the application of EEG brain mapping techniques offers a potential direction for further study of lying and self-deception. These techniques provide a more objective way to measure physiological responses and can help researchers better understand the cognitive processes underlying these phenomena.

5 Conclusion

The present study aimed to introduce a novel approach to polygraph deception scoring using a Multilayer Perceptron Neural Network (MLP) to predict high and low deception scores based on four psychophysiological indicators found to be highly predictive in the literature: amplitude of electrodermal reaction (ARED), amplitude of blood pressure in brachial pulse (ATAB), change of baseline level in chest breathing (MNBRT), and difference of altitude between breathing cycles (DIFA). The results of this study demonstrate that the MLP approach was able to correctly classify the deception level of the 400 offenders with an 80% correct classification rate (CCR), consistent with previous literature reporting similar prediction accuracy rates using machine learning approaches.

Our findings suggest that the use of psychophysiological indicators and machine learning algorithms represents a promising approach to identifying deception in polygraph examinations. One advantage of using an MLP over traditional polygraph scoring methods is its ability to simultaneously consider multiple psychophysiological indicators, which has been shown to improve detection accuracy. Moreover, using machine learning approaches such as MLP may help to minimize subjective scoring bias inherent in manual polygraph scoring methods.

It is important to acknowledge the various limitations of the current study. Initially, the research was based on a sample of offenders, which could limit its applicability to other groups. Subsequent investigations ought to strive to duplicate these results in more extensive and varied cohorts. Secondly, although the MLP technique demonstrated potential in precisely identifying dishonesty, it is crucial to acknowledge that it has certain constraints. The actual implementation of machine learning algorithms in real-world settings may be limited due to their high computational resource and expertise requirements. This study provides significant new understandings and opportunities for progress in the domains of law enforcement and forensic psychology. It also broadens the use of deep learning techniques to a new domain and contributes to the ongoing discourse on fraud detection.

The study's findings, which demonstrate an 80% accuracy rate in classifying deception scores, highlight the MLP approach's resilience in this situation and support the use of AI to improve the precision and dependability of conventional polygraph tests. However, future research should continue to investigate the reliability and validity of these approaches in larger and more diverse samples, and to explore ways to improve the practical applicability of these methods in real-world settings.

Funding

References

[1] Albrecht, W. S., Albrecht, C. O., Albrecht, C. C., Zimbelman, M. F. (2018). Fraud examination. Cengage Learning.

[2] Albu, R. D., Dzitac, I., Popentiu-Vlădicescu, F., Naghiu, I. M. (2014). Input Projection Algorithms Influence in Prediction and Optimization of QoS Accuracy. International Journal of Computers Communications & Control, 9(2), 131-138.

[3] Arasteh, A., Moradi, M. H., Janghorbani, A. (2016). A novel method based on empirical mode decomposition for P300-based detection of deception. IEEE Transactions on Information Forensics and Security, 11(11), 2584-2593.

[4] Ash, P. (1971). Screening employment applicants for attitudes toward theft. Journal of Applied Psychology, 55(2), 161.

[5] Baghel, N., Singh, D., Dutta, M. K., Burget, R., Myska, V. (2020, July). Truth identification from EEG signal by using convolution neural network: lie detection. In 2020 43rd International Conference on Telecommunications and Signal Processing (TSP) (pp. 550-553). IEEE.

[6] Baghel, N., Singh, D., Dutta, M. K., Burget, R., & Myska, V. (2020). Truth identification from EEG signal by using convolution neural network: lie detection. In 2020 43rd International Conference on Telecommunications and Signal Processing (TSP) (pp. 550-553). IEEE.

[7] Bashore, T. R., Rapp, P. E. (1993). Are there alternatives to traditional polygraph procedures? Psychological Bulletin, 113(1), 3–22. https://doi.org/10.1037/0033-2909.113.1.3

[8] Ben-Shakhar, G., Bar-Hillel, M., Lieblich, I. (1986). Trial by polygraph: Scientific and juridical issues in lie detection. Behavioral Sciences the Law, 4(4), 459-479.

[9] Bhutta, M. R., Hong, M. J., Kim, Y. H., Hong, K. S. (2015). Single-trial lie detection using a combined fNIRS-polygraph system. Frontiers in psychology, 6, 709.

[10] Bhuvaneswari, P., Kumar, J. S. (2015). A note on methods used for deception analysis and influence of thinking stimulus in deception detection. International Journal of Engineering and Technology, 7(1), 109-116.

[11] Brouwer, E. (2021). Schengen and the administration of exclusion: legal remedies caught in between entry bans, risk assessment and artificial intelligence. European Journal of Migration and Law, 23(4), 485-507.

[12] Cook, L. G., Mitschow, L. C. (2019). Beyond the polygraph: Deception detection and the autonomic nervous system. Federal Practitioner, 36(7), 316.

[13] Cook, L. G., Mitschow, L. C. (2019). Beyond the polygraph: Deception detection and the autonomic nervous system. Federal Practitioner, 36(7), 316.

[14] Cross, T. P., Saxe, L. (1993). A critique of the validity of polygraph testing in child sexual abuse cases. Journal of Child Sexual Abuse, 1(4), 19-34.

[15] Cunningham, M. R. (1989). Test-taking motivations and outcomes on a standardized measure of on-the-job integrity. Journal of Business and Psychology, 4(1), 119-127.

[16] Daneshi Kohan, M., Motie Nasrabadi, A., Shamsollahi, M. B., Sharifi, A. (2020). EEG/PPG effective connectivity fusion for analyzing deception in interview. Signal, Image and Video Processing, 14(5), 907-914.

[17] Dollins, A. B., Kraphol, D. J., Dutton, D. W. (2000). Computer algorithm comparison. Polygraph, 29(3), 237-257.

[18] Elkins, A. C., Gupte, A., Cameron, L. (2019, July). Humanoid robots as interviewers for automated credibility assessment. In International Conference on Human-Computer Interaction (pp. 316-325). Springer, Cham.

[19] Farrow, T. F., Burgess, J., Wilkinson, I. D., Hunter, M. D. (2015). Neural correlates of self-deception and impression-management. Neuropsychologia, 67, 159-174.

[20] Gamer, M. (2014). Mind reading using neuroimaging: Is this the future of deception detection?. European Psychologist, 19(3), 172.

[21] Gamer, M. (2014). Mind reading using neuroimaging: Is this the future of deception detection?. European Psychologist, 19(3), 172.

[22] Giansiracusa, N. (2021). Prevarication and the Polygraph. In How Algorithms Create and Prevent Fake News (pp. 99-118). Apress, Berkeley, CA.

[23] Gogate, M., Adeel, A., Hussain, A. (2017, November). Deep learning driven multimodal fusion for automated deception detection. In 2017 IEEE symposium series on computational intelligence (SSCI) (pp. 1-6). IEEE.

[24] Gordon, N. J. (2016). Essentials of polygraph and polygraph testing. CRC Press.

[25] Gordon, N. J., Mohamed, F. B., Faro, S. H., Platek, S. M., Ahmad, H., Williams, J. M. (2006). Integrated zone comparison polygraph technique accuracy with scoring algorithms. Physiology behavior, 87(2), 251-254.

[26] Happel, M. D. (2005). Neuroscience and the Detection of Deception. Review of Policy Research, 22(5), 667-685.

[27] Horvath, F. (2020). A Hundred Years of Polygraphy: Some Primary Changes and Related Issues. European Polygraph, 1, 31-45.

[28] Kohan, M. D., Nasrabadi, A. M., Shamsollahi, M. B. (2020). Interview based connectivity analysis of EEG in order to detect deception. Medical hypotheses, 136, 109517.

[29] Lakshan, I., Wickramasinghe, L., Disala, S., Chandrasegar, S., Haddela, P. S. (2019, October). Real time deception detection for criminal investigation. In 2019 National Information Technology Conference (NITC) (pp. 90-96). IEEE.

[30] Lapadusi, V., Dobreanu, R. (2014). History of Polygraph/Din Istoria Tehnicii Poligraf. Romanian Journal of Forensic Science, 15(95), 1737.

[31] Lo, Y. L., Fook-Chong, S., Tan, E. K. (2003). Increased cortical excitability in human deception. Neuroreport, 14(7), 1021-1024.

[32] Marouf, A. A., Ashrafi, A. F., Ahmed, T., Emon, T. (2019). A Machine Learning based Approach for Mapping Personality Traits and Perceived Stress Scale of Undergraduate Students. International Journal of Modern Education & Computer Science, 11(8).

[33] Marx, G. T. (1998). Ethics for the new surveillance. The Information Society, 14(3), 171-185.

[34] Masip, J. (2017). Deception detection: State of the art and future prospects. Psicothema, 29(2), 149-159.

[35] Matte, J. A. (1996). Forensic psychophysiology using the polygraph: Scientific truth verification, lie detection. JAM Publications.

[36] Matte, J. A. (2007). Critical Analysis of Krapohl's Validated Polygraph Techniques. Polygraph, 36(1), 29-34.

[37] Matte, J. A. (2020). Foremost Changes in Polygraph in Last 100 Years. European Polygraph, 14(1 (51)), 47-49.

[38] Meijer, E. H., Verschuere, B. (2017). Deception detection based on neuroimaging: Better than the polygraph?. Journal of Forensic Radiology and Imaging, 8, 17-21.

[39] Meijer, E. H., Verschuere, B. (2017). Deception detection based on neuroimaging: Better than the polygraph?. Journal of Forensic Radiology and Imaging, 8, 17-21.

[40] Nagle, D. E. (1983). The Polygraph in the Workplace. U. Rich. L. Rev., 18, 43.

[41] Nelson, R. (2015). Scientific basis for polygraph testing. Polygraph, 44(1), 28-61.

[42] Nelson, R. (2019). Literature Survey of Structural Weighting of Polygraph Signals: Why Double the EDA?. Polygraph & Forensic Credibility Assessment, 48(2), 105-112.

[43] Nevins, J. L. (2004). Measuring the mind: A comparison of personality testing to polygraph testing in the hiring process. Penn St. L. Rev., 109, 857.

[44] Nortje, A., Tredoux, C. (2019). How good are we at detecting deception? A review of current techniques and theories. South African Journal of Psychology, 49(4), 491-504.

[45] Nuwer, M. R. (1990). On the controversies about clinical use of EEG brain mapping. Brain Topography, 3(1), 103-111.

[46] Nuwer, M. R. (1990). The development of EEG brain mapping. Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society, 7(4), 459-471.

[47] O'Shea, J., Crockett, K., Khan, W., Kindynis, P., Antoniades, A., Boultadakis, G. (2018). Intelligent deception detection through machine based interviewing. In 2018 International joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

[48] Oravec, J. A. (2022). The emergence of "truth machines"?: Artificial intelligence approaches to lie detection. Ethics and Information Technology, 24(1), 1-10.

[49] Oswald, M. (2020). Technologies in the twilight zone: early lie detectors, machine learning and reformist legal realism. International Review of Law, Computers & Technology, 34(2), 214-231.

[50] Pasca, V. (2012). Study regarding psychophysiological reactivity values depending on subject's gender in polygraph testing. Procedia-Social and Behavioral Sciences, 33, 821-825.

[51] Pérez-Rosas, V., Abouelenien, M., Mihalcea, R., Burzo, M. (2015, November). Deception detection using real-life trial data. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 59-66).

[52] Popescu, M. C., Balas, V. E., Perescu-Popescu, L., Mastorakis, N. (2009). Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems, 8(7), 579-588.

[53] Rad, D., Magulod, G.C. Jr., Balas, E., Roman, A., Egerau, A., Maier, R., Ignat, S., Dughi, T., Balas, V., Demeter, E., Rad, G. and Chis, R. (2022). A Radial Basis Function Neural Network Approach to Predict Preschool Teachers' Technology Acceptance Behavior. Front. Psychol. 13:880753. doi: 10.3389/fpsyg.2022.880753.

[54] Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 19(4), 389.

[55] Raskin, D. C., Kircher, J. C. (2014). Validity of polygraph techniques and decision methods. In Credibility assessment (pp. 63-129). Academic Press.

[56] Raskin, D. C., Kircher, J. C., Honts, C. R. (2019). Computerized Polygraph Interpretations and Detection of Physical Countermeasures. In Anti-Terrorism; Forensic Science; Psychology in Police Investigations (pp. 179-188). Routledge.

[57] Raskin, D. C., Kircher, J. C., Honts, C. R., Horowitz, S. W. (2019). A Study of the Validity of Polygraph Examinations in Criminal Investigation: Final Report to the National Institute of Justice Grant No. 85-IJ-CX-0040. Polygraph, 48(1), 10-39.

[58] Saeed, S., Romarheim, A., Mancia, G., Saxvig, I. W., Gulati, S., Lehmann, S., Bjorvatn, B. (2022). Characteristics of hypertension and arterial stiffness in obstructive sleep apnea: A Scandinavian experience from a prospective study of 6408 normotensive and hypertensive patients. The Journal of Clinical Hypertension, 24(4), 385-394.

[59] Sánchez-Monedero, J., Dencik, L. (2022). The politics of deception borders:'biomarkers of deceit'and the case of iBorderCtrl. Information, Communication & Society, 25(3), 413-430.

[60] Selič, P. (2009). Analytical review of validated polygraph techniques. Journal of Criminal Investigation and Criminology/Ljubljana, 60, 4.

[61] Slowik, S. M. (2013). Evaluating Previously Conducted Polygraph Examinations. Polygraph, 42(4), 203-208.

[62] Slowik, S. M. (2020). Critical Changes Over the 100 Year Evolution of Polygraph Practices. European Polygraph, 14(1 (51)), 50-57.

[63] Twyman, N. W., Proudfoot, J. G., Schuetzler, R. M., Elkins, A. C., Derrick, D. C. (2015). Robustness of multiple indicators in automated screening systems for deception detection. Journal of Management Information Systems, 32(4), 215-245.

[64] Wahsheh, H., Al-Zahrani, M. (2021). Secure real-time computational intelligence system against malicious QR code links. International Journal of Computers Communications & Control, 16(3), 4186. https://doi.org/10.15837/ijccc.2021.3.4186

[65] Wahsheh, H., Al-Zahrani, M., ...Oprea, M., Mihalache, S. F., Popescu, M. (2017). Computational intelligence-based PM2.5 air pollution forecasting. International Journal of Computers Communications & Control, 12(3), 365-380.

[66] Walczyk, J. J., Igou, F. P., Dixon, A. P., Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. Frontiers in psychology, 4, 14.

[67] Wang, X., Shi, Y., Liu, L. (2021). Research on non-invasive psychological detection technology based on artificial intelligence. Academic Journal of Humanities & Social Sciences, 4(3).

[68] Widacki, J. (2020). A Half-Century of Experiences with the Polygraph. European Polygraph 14(1)51, 58-61.

[69] Zafran, E. L., Stickle, J. R. (1984). Polygraphs in Employment: A State Survey. Clev. St. L. Rev., 33, 751.

[70] Sinescu, C., Negrutiu, M., Pop, D., Cuc, L., DeSabata, A., Negru, R., Podoleanu, A. (2009, May). The importance of holograms in dentistry. In Holography: Advances and Modern Trends (Vol. 7358, pp. 138-147). SPIE.

[71] Toader, C. S., Rujescu, C. I., Feher, A., Sălăşan, C., Cuc, L. D., Bodnár, K. (2023). Generation differences in the behaviour of household consumers in Romania related to voluntary measures to reduce electric energy consumption. Amfiteatru Economic, 25(64), 710-727.

Cite this paper as:

Rad, D.; Kiss C.; Rad, G.; Paraschiv, N.; Bălaş, M. (2025). A multilayer perceptron neural network prediction approach to polygraph scoring , *International Journal of Computers Communications & Control*, 20(2), 7008, 2025.
https://doi.org/10.15837/ijccc.2025.2.7008