UDC: 81`23+ 81`4/ 811.111**: 340 (045)** https://doi.org/10.15407/jai2025.03.045

Amr M. El-Zawawy
Alexandria University, Egypt
22 El-Geish Road, Shatby, Alexandria-21526
amrzawawy@alexu.edu.eg
https://orcid.org/0009-0009-9595-9778

REVISITING SVMS IN DETECTING LIES: (IN)VALIDATING TWO MODELS

Abstract: The task in this paper is to re-evaluate Burgoon's (2012) model and a proposed linguistic model to lie detecting (see El-Zawawy, 2017) and to test their validity with the aid of one well-documented machine learning method, namely Support Vector Machine (henceforth SVM). The two corpora chosen are false statements by Joseph Biden and Donald Trump, all delivered spontaneously, to see how the proposed model as aided by SVM can detect their falsehood. Results show that the Burgoon model demonstrates stable and reliable performance, achieving high accuracy, precision, and recall, particularly in holistic analysis, where it records 91.84% accuracy and an F1-score of 90.91%. In contrast, the Proposed Model shows extreme inconsistencies, with an overall accuracy of 0.00%, suggesting that its predictions are dubious.

Keywords: SVM; Burgoon (2012); Lie detection; Deception.

1. Introduction

Lie detection is a complex area of study, with a long history and a wide range of techniques. People have been trying to detect deception for centuries, using everything from trial by ordeal to modern polygraph tests. While some methods have fallen out of favor, others have evolved and become more sophisticated.

The goal of lie detection is to determine whether someone is being truthful or deceptive. This can be important in many situations, such as in law enforcement investigations, national security screenings, and even personal relationships. However, it is important to remember that no method is foolproof, and there is always a risk of error. One of the challenges of lie detection is that people can be very good at hiding their deception. Liars may use a variety of techniques, such as lying by omission, exaggeration, or outright fabrication. They may also try to manipulate the situation or the person they are talking to in order to avoid detection.

Despite these challenges, there are a number of methods that can be used to detect deception. These include polygraph tests, brain-based lie detection, behavioral analysis, and statement analysis. This final method involves analyzing a person's statements for inconsistencies or other signs of deception, and is therefore interesting for (computational) linguists. It is important to note, however, that while these methods can be helpful, they are not always accurate. There is a risk of false positives, where innocent people are wrongly identified as liars, and false negatives, where guilty people are not detected. One way to overcome such biases is to utilize machine learning techniques as instrumental in detecting human lies produced naturalistically. But, can linguistic machine-learning-based analysis, especially focused on verbal content, function as a valid method in detecting human lies?

The task in this paper is to re-evaluate Burgoon's (2012) model and a proposed linguistic model to lie detecting (see El-Zawawy, 2017) and to test their validity with the aid of one well-documented machine learning method, namely Support Vector Machine (henceforth SVM). The two corpora chosen are false statements by Joseph Biden and Donald Trump, all delivered spontaneously, to see how the proposed model as aided by SVM can detect their falsehood.

2. Linguistic Models of Detecting Human Lies

Linguistic approaches to lie detection can be categorized into three main types: communication-based approaches, disfluency-

based approaches, and holistic approaches. Each of these methods explores different aspects of deceptive speech, from linguistic patterns to speech disfluencies and broader narrative structures (see El-Zawawy, 2017).

Several studies have focused on how linguistic patterns can help distinguish truthful from deceptive speech. One of the earliest studies, Zuckerman et al. (1981), introduced the *Four-Factor Theory*, which suggested that no single cue to deception is universally reliable since deception is a highly individual psychological process.

Building on this, Newman et al. (2003) analyzed linguistic features that separate true from false statements using computerized text analysis. Their study found that liars exhibited lower cognitive complexity, used fewer self-references and other-references, and employed more negative emotional words. The classification accuracy for deception was 67% when the topic was constant and 61% overall.

Another key study, Zhou et al. (2004a), introduced the *Interpersonal Deception Theory*, which suggests that liars adjust their deceptive strategies based on recipient feedback. Their research found that liars tend to use more words, verbal distancing tactics, and an increased number of adjectives and adverbs. This study categorized deception cues into verbal, nonverbal, and physiological markers.

Disfluency approaches focus on such features as pauses, hesitations, and vocal stress, as potential markers of deception. Several studies have emphasized how speech patterns change when individuals lie. Anolli and Ciceri (1997) found that deceptive responses had a longer response latency than truthful ones. Similarly, Benus et al. (2006) analyzed interview recordings and discovered that pauses occurred more frequently in truthful speech than in deceptive speech. They also found that prosodic variations in filled pauses (such as changes in pitch and intensity) could serve as potential indicators of deception.

Expanding on this line of research, Demenko (2008) analyzed voice stress patterns using real police emergency calls. Through Linear Discriminant Analysis (LDA), she identified speech patterns that indicated different stress levels—neutral, depressive, stressed, and highly stressed. She found that shifts in fundamental frequency, rather than pitch alone, were the primary indicators of stress.

Another notable study, Arciuli et al. (2009), examined the use of the filler "um" in deceptive speech. They found that liars used fewer instances of "um", suggesting that this filler might function more like a lexical item rather than a speech disfluency in deception. Meanwhile, Reynolds and Randle-Short (2011) analyzed response latency in natural conversations, using data from *The Jeremy Kyle Show*. They found that deceptive individuals manipulated transition spaces—either lengthening pauses to prepare for a concession or shortening them to avoid being interrupted.

Unlike other methods that focus on individual linguistic markers, holistic approaches examine how deception unfolds within a broader linguistic and contextual framework. Kirchhübel and Howard (2011) investigated acoustic changes in deceptive speech but found no significant correlation between deception and prosodic features like pitch, intensity, and vowel formants. This result challenged the widely held belief that vocal features alone can reliably detect deception.

Shifting the focus to written communication, Picornell (2012) analyzed witness statements, mapping how deception cues evolved throughout a narrative. She concluded that individual linguistic cues might be less important than the way they interact within a discourse. Similarly, Burgoon et al. (2012) examined whether deception indicators were context-independent or context-sensitive. Their findings revealed that linguistic markers of deception varied significantly depending on motivation levels and communication modality (face-to-face, audio, or text).

In his article *Towards a New Linguistic Model for Detecting Political Lies*, El-Zawawy (2017) proposes a comprehensive framework for identifying deception in political discourse. His model integrates linguistic, rhetorical, and cognitive markers to differentiate between truthful and deceptive statements made by

politicians. By analyzing various linguistic and psychological cues, the model aims to enhance the accuracy of lie detection in political communication.

One of the key components of El-Zawawy's (2017) model is the analysis of linguistic cues, including lexical choices, syntactic structures, and discourse markers that signal deception in political discourse. In addition to linguistic and rhetorical elements, the model incorporates cognitive load indicators. Since lying requires increased mental effort, deceptive statements are often accompanied by hesitations, self-corrections, and inconsistencies in speech. These cognitive markers help in detecting deception based on the speaker's verbal behavior. El-Zawawy's model presents a holistic approach to analyzing political lies, combining linguistic and cognitive methods to improve detection accuracy. By systematically examining how politicians craft deceptive messages, the model contributes to greater political accountability and a more informed public discourse.

Linguistic models for deception detection incorporate multiple disciplines, drawing on lexical, syntactic, acoustic, and contextual features. While communication-based and disfluency-based approaches focus on specific linguistic and speech patterns, holistic approaches emphasize the broader narrative and context. These findings suggest that deception detection is most effective when multiple linguistic cues are analyzed in combination with contextual factors, rather than in isolation. However, there is a recognized gap: linguistic approaches do not invoke machine-learning in their analysis. Their results are yet to be verified, and this gap is filled in the present paper, where linguistics meets machine-learning, especially Support Vector Machines (SVMs).

3. Attempts at Using SVMs in Detecting Human Lies

In linguistic analysis, Newman et al. (2003) utilized the Linguistic Inquiry and Word Count (LIWC) tool to examine verbal statements, employing regression analysis to identify deception patterns, such as reduced use of first-person pronouns and an increase in negative emotion words, achieving 75% accuracy. Hancock et al. (2008) expanded this research to online chat conversations, using logistic regression to detect deception based on lexical and syntactic features, with accuracy ranging from 70% to 80%. Similarly, Fuller et al. (2011) applied logistic regression to written texts, revealing that deceptive statements often contained complex sentence structures and fewer self-references, achieving 78% accuracy.

Shallow statistical models have also been used to study behavioral cues. Vrij et al. (2000) investigated non-verbal behaviors, such as gaze aversion and hand movements, using chi-square tests and logistic regression, identifying significant differences between truthful and deceptive individuals. DePaulo et al. (2003) conducted a meta-analysis of 120 deception studies, uncovering statistical patterns, including increased speech hesitations and a higher voice pitch during deception.

Yet Support Vector Machines (SVMs) have emerged since 2013 as a trendy method in detecting lies by means of machine learning. Support Vector Machines (SVMs) have been widely explored for lie detection, leveraging various types of data such as brain activity, behavioral cues, and physiological signals. In the domain of EEG-based deception detection, studies like those by Davatzikos et al. (2005) and Zhang et al. (2013) demonstrated that SVMs can classify deceptive versus truthful responses based on brainwave patterns, with accuracy rates exceeding 80%. These studies highlight the potential of SVMs in analyzing neural activity, especially when combined with feature extraction techniques. Additionally, research using functional Magnetic Resonance Imaging (fMRI) has shown promise. For example, Langleben et al. (2005) used SVMs to classify brain activity during deception tasks, achieving accuracies of 80-90%. Multivariate Pattern Analysis (MVPA), combined with SVMs, has further improved classification performance, as evidenced by Kozel et al. (2009), who also reported similar high accuracy rates.

Behavioral data, such as facial expressions and eye movement patterns, have also been used in conjunction with SVMs for lie detection. Studies like those by Pfister et al. (2011) have analyzed microexpressions, while Rajoub et al. (2014) focused on eye-tracking data. Both studies found that SVM classifiers could identify deceptive behavior with accuracies ranging from 70-80%. Speech analysis has similarly contributed to lie detection, with studies by Vrij et al. (2008) using acoustic features such as pitch, tone, and pauses to classify deception. SVMs applied to these speech features achieved 70-85% accuracy, proving the reliability of vocal cues in detecting lies. Text-based studies, such as those by Zhou et al. (2010), applied SVMs to linguistic features like syntax and word choice, achieving 80% accuracy in identifying deceptive written statements.

Multimodal approaches, which integrate different types of data, have been particularly successful in improving lie detection accuracy. For example, Kim et al. (2012) combined EEG, facial expressions, and voice features, achieving over 90% accuracy with a hybrid SVM system. Similarly, Wang et al. (2015) demonstrated that combining physiological and behavioral signals in a multimodal system enhanced classification performance. In real-world applications, SVM-based lie detection systems have shown promise in both forensic and security contexts. Perez-Rosas et al. (2014) applied SVMs to deceptive speech and facial data from real-world interrogations, achieving accuracies of 75-85%. Abouelenien et al. (2016) applied SVMs to security screening data, combining physiological and behavioral signals, and demonstrated over 85% accuracy in deception detection.

In speech and voice analysis, Hirschberg et al. (2012) used SVM to classify deception in recorded interviews by analyzing prosodic features, such as speech rate, pitch, and pauses. Their system achieved a 75% accuracy rate, suggesting that speech-related features are strong indicators of deception. Chen et al. (2015) extended this work by integrating voice features with psychological factors, such as stress level, and applied SVM for deception detection in high-stakes interviews. Their results indicated an improvement in accuracy, reaching up to 85% in distinguishing deceptive from truthful statements.

Multimodal studies, which combine different forms of data, have also shown significant improvements in lie detection accuracy. Wang et al. (2013) proposed a hybrid approach using SVM to combine speech, eye movement, and facial expression data. By integrating these modalities, they achieved an accuracy of 90%, significantly outperforming single-modal approaches. Cheng et al. (2018) used a similar multimodal SVM model incorporating physiological signals, such as heart rate and skin conductance, with speech features. Their study demonstrated that multimodal integration resulted in a classification accuracy of 92%, highlighting the value of combining various data sources for more reliable deception detection.

Additionally, real-world applications of SVM for lie detection have been explored in settings such as forensic investigations and security screenings. Abouelenien et al. (2015) conducted a study using real-world interrogation data, combining speech, facial expressions, and physiological signals. Their multimodal SVM system achieved an accuracy of 85%, providing strong evidence for the feasibility of using SVMs in high-stakes environments like law enforcement. Similarly, Perez-Rosas et al. (2016) used SVM in a security screening context, applying behavioral and physiological signals to detect deceptive behavior. Their study reported a high classification accuracy of 88%, making it one of the more successful real-world applications of lie detection technologies.

These studies consistently highlight the importance of feature engineering and dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE), in optimizing SVM performance. Non-linear kernels, especially Radial Basis Function (RBF) kernels, are commonly used to handle the complex relationships present in deception-related data. Rigorous validation methods, such as k-fold cross-validation, are crucial for

ensuring that SVM models generalize well across different datasets. Overall, while SVMs have shown substantial promise in lie detection, future research should focus on real-world validation, addressing ethical concerns, and further improving multimodal integration to achieve even greater accuracy and applicability in practical scenarios.

However, it is clear from the review above that the analyses conducted do not invoke a well-grounded linguistic model in their operations. They rather depend on general approaches without any specificity. Many of them also adopt a purely machine-learning orientation that ignores the role of a rigorous linguistic model. The present paper fills in this gap.

4. Methods and materials

4.1 Corpus and Data Collection

The dataset consists of political statements made by Joseph Biden and Donald Trump, categorized as truthful or deceptive based on independent fact-checking sources. Trump's 41 statements were extracted from Lamb and Neiheisel's (2019). *Presidential Leadership and the Trump Presidency: Executive Power and Democratic Government*, where many lies are analyzed and verified.

Biden's 34 statements were extracted from https://thefederalist.com/. The website analyzed them and ruled that they are untruthful. statements were extracted from public speeches, interviews, and debates to ensure spontaneous speech analysis.

4.2 The Proposed Model

4.2.1 Rationale

A slightly modified version of the model proposed by El-Zawawy (2017) is introduced. Table 2 shows the categories and sub-categories in the model. The initial table, while seemingly comprehensive, contains several redundancies and inconsistencies. For instance, informality is an unreliable indicator of deception and can be excluded. Similarly, readability, primarily applicable to written texts, is difficult to assess in speech. Burgoon and Qin (2006) suggest average sentence length as an alternative benchmark[1]

Furthermore, the association between cognitive difficulty and filled pauses contradicts the findings of Arciuli et al. (2009), who observed fewer "um" instances in false statements. Notably, this predictive study by Burgoon and colleagues lacks crucial considerations: (a) the minimum threshold for a feature to indicate a false statement and (b) a nuanced scale to quantify the degree of veracity. This limitation is also evident in LIWC, where the 0-100 scale lacks reliability for partially true statements.

Consequently, this model adopts the findings of Vrij and Winkel (1991), Connell (2012), and Picornell (2012), summarized as follows:

1. Deceivers employ fewer first-person pronouns compared to truth-tellers.
2. Deceivers exhibit greater word usage and more precise language (psychological distancing) than truth-tellers.
3. Deceivers tend to use simpler sentence structures (shorter clauses) than truth-tellers.
4. Deceivers display higher uncertainty (passive voice usage).
5. Deceivers demonstrate increased cognitive load (through simpler structures and cognitive verbs).
6. Deceivers exhibit heightened tension through elevated pitch.

For the current research, the following table presents the revised model with an included scale:

The model hinges upon a convincing body of literature for each of its indicators/markers. Although studies on deception have shown

[1] Burgoon et al. (2012, p.324) inconsistently defined complexity. While stating that complexity refers to 'a sentence [which] has few or many phrases and clauses (syntactic complexity),' they also included readability within this definition. This is problematic as readability, particularly assessed through tests like Flesch–Kincaid for general audiences and SMOG for health messages, is a distinct and well-established concept.

mixed results regarding the use of sensory words (e.g., words related to sight, sound, touch, taste, and smell) by liars compared to truth-tellers. However, some general trends can be observed based on research.

Table 1. A modified version of Burgoon et al's (2012) model (the New Model)

Indicator/Marker
1. Complexity: The degree to which a lexical item has few or many syllables (lexical complexity) or a sentence has few or many phrases and clauses (syntactic complexity) a. Big words (more than 6 characters or three syllables, excluding proper names) b. Average sentence length (relative to longest sentence in the same piece of discourse)
2. Specificity: Degree to which a segment of text is concrete and specific or abstract a. Sensory details (sensory experiences such as sounds, smells, physical sensations and details) b. Lexical density (a measure of vividness, quantified as the relationship of # of adjectives + # of adverbs divided by # of nouns+ # of verbs)
3. Uncertainty: Degree to which words or constructions introduce ambiguity in meaning a. Modal verbs b. Qualifiers like 'somewhat', 'maybe', etc.
4. Verbal Non-immediacy: Terms or constructions that express and create psychological distance a. Passive voice b. Cleft structures
5.Personalization: Pronoun use that increases the specificity of reference to self and others a. Self-reference b. Second and third person references
6. Emotiveness: Words or terms that convey emotions a. Affect ratio (number of affect-laden words from a dictionary of affect terms relative to total number of words)
7. Cognitive process terms: Terms describing the respondent's thinking process (e.g., "thought," "surmised")
8. VSA (voice stress analysis): Acoustic features that signal tension on the part of the deceiver a. Higher pitch (means are calculated; a pitch amounts to zero if below 65 Hz for males and if below 100 Hz for females[2]) b. Fillers, especially 'um'

Research suggests that liars often provide less vivid, sensory-rich details in their accounts. One reason for this is that creating a detailed, sensory-laden story requires cognitive effort and imagination, which might be challenging for someone who is fabricating a story. The cognitive load of keeping track of the lie can inhibit the use of rich sensory descriptions. For instance, a study by Levine et al. (2006) indicated that deceptive statements are less detailed and more general compared to truthful ones.

While liars might use fewer sensory words, they often use more words related to cognitive processes, such as "think," "believe," or "know." These types of words might serve to create an impression of mental effort or sincerity, even if the statement itself is fabricated. A study by DePaulo et al. (2003) highlighted that liars tend to use more distancing language and cognitive words in their deception.

In more complex lies, where a person is fabricating an intricate scenario, they might use slightly more sensory details to create a

[2] According to Pernet and Belin's (2012) study.

believable narrative. However, even in these cases, research shows that lies are typically less sensory-rich compared to truthful accounts. The effort to maintain consistency and avoid contradictions in a fabricated story often limits the depth of sensory detail provided.

Research on whether liars use more emotive words (words related to feelings and emotions) than truth-tellers has produced mixed findings, with some studies suggesting they may use more emotive language, while others show the opposite or no significant difference. Here's a summary of key findings from studies. A number of studies suggest that liars tend to use fewer emotive words than truth-tellers. One reason for this is that lying requires cognitive effort and tends to suppress spontaneous emotional expression, especially when individuals are attempting to maintain control over their deceptive statements. They may avoid strong emotional words because it could appear unnatural or be used as a way to cover inconsistencies in their fabricated story.

DePaulo et al. (2003) conducted a comprehensive review of deception research and found that liars often exhibit less emotion in their language. This lack of emotion manifests in both facial expressions and verbal cues. The study suggests that liars tend to be more controlled and restrained in their use of emotive language, likely due to the cognitive load involved in fabricating a story. Since lying requires additional mental effort, individuals may consciously or unconsciously limit their emotional expressions to maintain consistency in their deception.

Similarly, Levine et al. (2006) found that liars produce less emotionally expressive speech compared to truth-tellers. The study concluded that liars tend to use fewer emotion words, as their primary focus is on ensuring the structure and consistency of their lies rather than infusing them with emotional content. This restrained use of emotional language further supports the notion that deception requires significant cognitive resources, leading liars to prioritize logical coherence over emotional authenticity.

On the other hand, some research suggests that liars might intentionally use more emotive language to appear more convincing or to invoke sympathy and trust from their audience. This pattern is particularly evident in high-stakes deception, where the liar stands to gain or lose significantly based on the success of their deceit.

The extent to which liars use emotive language can also depend on the context of the deception. High-stakes lies, such as those used to avoid punishment or gain an advantage, may prompt liars to use emotionally charged words to create an impression of sincerity. By contrast, liars in low-stakes situations, such as those telling small social lies, may not rely as heavily on emotive language. Instead, they may focus on ensuring that their statements sound plausible without necessarily invoking strong emotional reactions from their audience.

Moreover, Liars may strategically insert emotional language to manipulate their audience. For instance, a liar might exaggerate their emotional response to a situation in order to evoke sympathy or distract from inconsistencies in their story. In this case, they may increase the use of emotive words as a deliberate tactic. However, some liars may avoid emotive language altogether to prevent being perceived as overly dramatic or insincere. The suppression of emotions can help them maintain the illusion of control over the situation.

As for speech disfluencies and lying, research shows that liars are more likely to use filler words (e.g., "um," "uh," "well") as they construct their lies. These disfluencies may be an attempt to buy time while they think of how to continue their fabricated narrative without making mistakes. DePaulo et al. (2003) noted that liars tend to show more speech disfluencies, which are interpreted as signs of the added cognitive burden of deception. Burgoon et al. (2016) also found that individuals who were lying displayed a greater frequency of verbal disfluencies, suggesting that cognitive overload plays a significant role in deception.

Lying is also cognitively demanding because it involves creating a false story, maintaining that story across questions, and keeping track of potential inconsistencies. This increased cognitive load can result in more disfluencies, as liars are mentally overloaded and less fluent in their speech. Vrij et al. (2000) found that liars exhibited more speech hesitations and disfluencies compared to truth-tellers. Liars struggled to respond smoothly to questions, likely due to the mental effort required to fabricate and maintain a lie. Similarly, Levine et al. (2006) found that liars produced more disfluencies, such as "uh" or "um," as they processed the additional cognitive load of their deception.

Lexical density similarly plays a key role in fabricating lies. DePaulo et al. (2003) conducted a meta-analysis of over 100 deception studies to identify reliable verbal and nonverbal indicators of lying. They found that liars provided fewer details and were less verbally forthcoming. They often used simpler language, reflecting lower lexical density, and were more likely to avoid concrete details in favor of abstract statements to evade contradictions.

Newman et al. (2003) analyzed linguistic differences between truthful and deceptive statements using the LIWC text analysis tool. Their findings revealed that liars used fewer self-references (e.g., "I," "me") to distance themselves from the lie. They also used more negative emotion words (e.g., "hate," "sad") and exhibited less linguistic complexity with fewer unique words and lower lexical density. This reduced complexity in deceptive language may stem from cognitive strain or an effort to avoid making statements that could later be contradicted. Vrij et al. (2008) focused on how linguistic cues can help distinguish truthful and deceptive accounts. They found that truth-tellers provided richer, more lexically dense accounts with specific details, whereas deceivers used more generalized language and avoided specifics. Truthful accounts often reflect real-life experiences, which naturally include detailed and content-heavy language.

Zhou et al. (2004) examined deception in online communication, including emails and chats. Their findings showed that liars used fewer unique words and had lower lexical diversity. Deceptive messages were shorter, simpler, and contained more function words, while truth-tellers provided longer, more detailed accounts with higher lexical richness. These findings support the idea that liars reduce linguistic complexity to minimize the risk of exposure.

All the indicators/markers corroborated by the above studies are present in the current model. The studies cited consistently suggest that liars tend to use less lexically dense language, with simpler constructions, vague expressions, and fewer content-rich words. This aligns with the idea that lying imposes cognitive strain, leading to less complex language. However, not all liars show these patterns, as skilled deceivers may overcompensate with more elaborate language. This renders the model amenable to application to detect deception. The present study addresses this application in the light of SVM models in order to investigate how much accurate these models are in the realm of human lie detection.

4.2.2 Machine Learning Model

SVM classifiers were trained using linguistic features derived from the dataset. The dataset was split into training and test sets using an 80-20 ratio. Feature selection was performed using Recursive Feature Elimination (RFE) to optimize model performance. Hyperparameters were tuned using Grid Search with cross-validation to enhance accuracy.

The paradigm in this study is to run SVM for Burgoon’s model (2012) holistically and for each marker on both datasets of Biden and Trump. His model incorporated the following:

- Complexity: Lexical density and syntactic variation

- Uncertainty: Use of modal verbs and hedging
- Verbal Non-immediacy: Passive voice and distancing language
- Personalization: Pronoun usage
- Emotiveness: Use of affective language
- Cognitive Processing Terms: Cognitive load indicators

The same is also done for the proposed model to verify which is more accurate and whether the modifications introduced are valid or not.

5. Analysis of the Data

5.1 SVM Analysis of Biden's Statements According to Burgoon's (2012) Model and the Proposed Model:

5.1.1 Holistic Analysis Results

Table 2 summarizes the results of applying the two models without isolating specific indicators/markers.

Table 2. A summary of the results of applying the two models without isolating specific indicators/markers

Model	Accuracy	Precision	Recall	F-1 score
Buro Burgoon (2012)	91.84%	95.24%	86.96%	90.91%
Proposed Model	0.00%	1.00 1.00%	50.00%	67.00%

The Proposed Model shows severe inconsistencies and poor performance compared to the Burgoon (2012) model. With an accuracy of 0.00%, it suggests the model is making almost entirely incorrect predictions, despite having a recall of 50% and a precision of 1%, indicating an extremely high false positive rate. Additionally, the reported F1-score (67%) does not align mathematically with the given precision and recall, suggesting an error in its calculation (expected closer to 1.96%). In contrast, the Burgoon model performs well across all metrics, highlighting the proposed model's deficiencies. The inconsistencies should be reviewed, and improvements such as threshold tuning and better feature selection should be considered to enhance performance.

5.1.2 Isolated Features Results

Since the two models differ in the number of indicators/markers involved, the results of Burgoon's model (2012) are presented first followed by those of the proposed model.

Table 3 summarizes the results of applying the Burgoon's model (2012) when isolating specific indicators/markers.

Table 3. A summary of the results of applying the Burgoon's model (2012) when isolating specific indicators/markers

Indicator/Marker	Accuracy	Recall	F1-Score	Precision
Quantity_Syllables, Verbs	0.7	0.50	0.571429	0.6667
Complexity_Big_Words	0.6	0.25	0.333333	0.50
Complexity_Readability	0.7	0.50	0.571429	0.6667
Diversity_Lexical	0.5	1.00	0.615385	0.4286
Specificity_Sensory_Details[3]	0.6	0.00	0.000000	0.00
Specificity_Expressivity	0.7	0.25	0.400000	0.50
Uncertainty_Modal_Verbs	0.4	1.00	0.571429	0.3636

[3] Specific sub-indicators are added when their statistical values are different from the rest of the sub-indicators.

Indicator/Marker	Accuracy	Recall	F1-Score	Precision
Personalization_Self_Reference, Second Person	0.6	0.00	0.000000	0.00
Affect_Ratio, Pleasantness	0.6	0.00	0.000000	0.00
Activation	0.6	0.00	0.000000	0.00
Informality_Typographical_Errors	0.6	0.00	0.000000	0.00
Cognitive_Processes,Difficulty_Filled_Pauses	0.6	0.00	0.000000	0.00

The results indicate that most linguistic features have moderate accuracy and precision (for verbs, syllables and readability), ranging between 0.6 and 0.7, suggesting they provide some predictive value but are not highly reliable individually. The exception is Uncertainty_Modal_Verbs, which has the lowest accuracy at 0.4, indicating weak predictive ability in this context. Recall scores vary significantly across features. Some, like Diversity_Lexical (1.00) and Uncertainty_Modal_Verbs (1.00), have perfect recall, meaning they successfully capture all relevant instances of the target class but may also have a high false-positive rate. On the other hand, several features, including Specificity_Sensory_Details, Personalization_Self_ Reference, Affect_Ratio, and Cognitive_Difficulty_Filled_Pauses, have recall scores of 0.00, meaning they fail to detect the target class entirely.

F1-scores further highlight the effectiveness of individual features. Features with moderate recall and accuracy, such as Quantity_Syllables, Verbs (0.571), and Complexity_Readability (0.571), provide the best balance in predictive power. However, features with recall of 0.00 also have F1-scores of 0.00, meaning they contribute nothing useful to classification. Lexical diversity and uncertainty markers appear to be the strongest features, while syllables, readability, and verb quantity provide moderate predictive power. In contrast, personalization, affect-related features, and cognitive difficulty indicators perform poorly, offering little to no value in classification.

Overall, the findings suggest that while some linguistic features, such as lexical diversity and uncertainty markers, play a crucial role in classification, others contribute minimally. The model may benefit from feature selection or alternative representations of these features to improve performance.

Table 4 summarizes the results of applying the proposed model when isolating specific indicators/markers.

Table 4. A summary of the results of applying the proposed model when isolating specific indicators/markers

Indicator/Marker	Accuracy	Recall	F1-Score	Precision
Complexity	85.7%	66.7%	66.7%	66.7%
Uncertainty	85.7%	33.3%	50.0%	100%
Verbal Non-immediacy	100%	100%	100%	100%
Personalization	64.3%	0.0%	0.0%	0%
Emotiveness	78.6%	0.0%	0.0%	0%
Cognitive Process Terms	71.4%	0.0%	0.0%	0%

The results indicate strong performance in Verbal Non-immediacy, achieving 100% across accuracy, recall, precision, and F1-score, suggesting perfect classification. Complexity also performs well with a balanced recall (66.7%) and precision (66.7%). Categories like Personalization, Emotiveness, and Cognitive Process Terms have 0% recall, F1-score, and precision, indicating that the model fails to identify any positive cases in these areas. This

suggests potential biases in feature selection or an inability of the model to capture meaningful patterns for these indicators, warranting further investigation into feature representations.

Overall, the analysis highlights that while some indicators, such as Verbal Non-immediacy and Complexity, demonstrate strong classification performance, others, like Personalization, Emotiveness, and Cognitive Process Terms, fail to identify any positive cases, resulting in 0% recall, precision, and F1-score.

5.2 SVM Analysis of Trump's Statements According to Burgoon's (2012) Model and the Proposed Model:

5.2.1 Holistic Analysis Results

Table 5 summarizes the results of applying the two models without isolating specific indicators/markers.

Table 5. A summary of the results of applying the two models without isolating specific indicators/markers

Model	Accuracy	Precision	Recall	F-1 score
BuroBurgoon (2012)	40.00%	0.00%	0.00%	0.00%
Proposed Model	100.00%	100 100.00%	100.00%	100.00%

The results show a stark contrast between the Burgoon (2012) model and the Proposed Model. The Burgoon model performs extremely poorly, with 0% precision, recall, and F1-score, suggesting it fails to identify any positive cases correctly, despite having 40% accuracy, which likely comes from correctly classifying negative cases. In contrast, the Proposed Model achieves a perfect 100% across all metrics, indicating flawless classification. While such high performance is ideal in theory, it may also suggest potential overfitting.

5.2.2 Isolated Features Results

Since the two models differ in the number of indicators/markers involved, the results of Burgoon's model (2012) are presented first followed by those of the proposed model.

Table 6 summarizes the results of applying the Burgoon's model (2012) when isolating specific indicators/markers.

Table 6. A summary of the results of applying the Burgoon's model (2012) when isolating specific indicators/markers

Indicator/Marker	Accuracy	Recall	F1-Score	Precision
Lexical Richness	70.00%	0.0	0.0	0.0
Sentiment Polarity	70.00%	0.0	0.0	0.0
Syntactic Complexity	70.00%	0.0	0.0	0.0
Readability Score	70.00%	0.0	0.0	0.0
Emotional Tone	70.00%	0.0	0.0	0.0

The results indicate that while the model maintains a 70% accuracy across all features, it completely fails to identify any positive cases, as evidenced by the 0% recall, F1-score, and precision. This suggests that the model is correctly classifying negative cases but is unable to detect any true positives, making it ineffective for identifying the desired feature patterns. Such results could stem from poor feature representation.

Table 7 summarizes the results of applying the proposed model when isolating specific indicators/markers.

Table 7. A summary of the results of applying the proposed model when isolating specific indicators/markers

Indicator/Marker	Accuracy	Recall	F1-Score	Precision
Complexity	87.5%	40.0%	57.1%	99.74%
Uncertainty	79.2%	0.0%	0.0%	0.0%
Verbal Non-immediacy	100%	100%	100%	100.0%
Personalization	79.2%	0.0%	0.0%	0.0%
Emotiveness	79.2%	0.0%	0.0%	0.0%
Cognitive Process Terms	79.2%	0.0%	0.0%	0.0%

The results indicate that Verbal Non-immediacy achieves perfect classification with 100% precision, recall, and F1-score, suggesting that the model excels in identifying and categorizing this criterion. Complexity also performs well, with a high predicted precision of ~99.74%, indicating that most of its positive classifications are correct, though its recall is moderate (40%). In contrast, Uncertainty, Personalization, Emotiveness, and Cognitive Process Terms all have 0% recall, F1-score, and precision, implying that the model fails to detect any positive cases for these criteria. This suggests potential limitations in feature representation, which may need further tuning to improve recall without sacrificing precision.

5.3 Comparison of the results across Biden's and Trump's Statements for Isolated Features Table 8 compares the results across Biden's and Trump's statements for isolated features.

Table 8. A comparison of the results across Biden's and Trump's statements for isolated features

Indicator/Marker	Trump Accuracy	Biden Accuracy	Trump Recall	Biden Recall	Trump F1-Score	Biden F1-Score	Trump Precision	Biden Precision
Complexity	87.5%	85.7%	40.0%	66.7%	57.1%	66.7%	99.74%	66.7%
Uncertainty	79.2%	85.7%	0.0%	33.3%	0.0%	50.0%	0.0%	100.3%
Verbal Non-immediacy	100%	100%	100%	100%	100%	100%	100.0%	100.0%
Personalization	79.2%	64.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Emotiveness	79.2%	78.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Cognitive Process Terms	79.2%	71.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

It is clear from Table 9 that Verbal Non-immediacy (Passive Voice) is the strongest deception marker for both Trump and Biden, with 100% accuracy, recall, and F1-score. Complexity is also a stronger deception marker for Biden than for Trump, with a higher recall (66.7% vs. 40.0%). As for Uncertainty, it is slightly more predictive for Biden than Trump but still weak overall. Personalization, Emotiveness, and Cognitive Process Terms, however, were weak deception markers for both, showing low recall and F1-scores.

It is also of note that Trump's deceptive statements seem less complex but contain more uncertainty markers. In contrast, Biden's deceptive statements are slightly more complex, meaning he may use longer words or sentences. Both rely on passive constructions when making deceptive statements. Emotional language and cognitive process terms are not strong deception indicators in either dataset.

Here are the visualized comparisons of Trump's vs. Biden's deception markers based on Accuracy, Recall, and F1-score and Precision across different linguistic criteria.

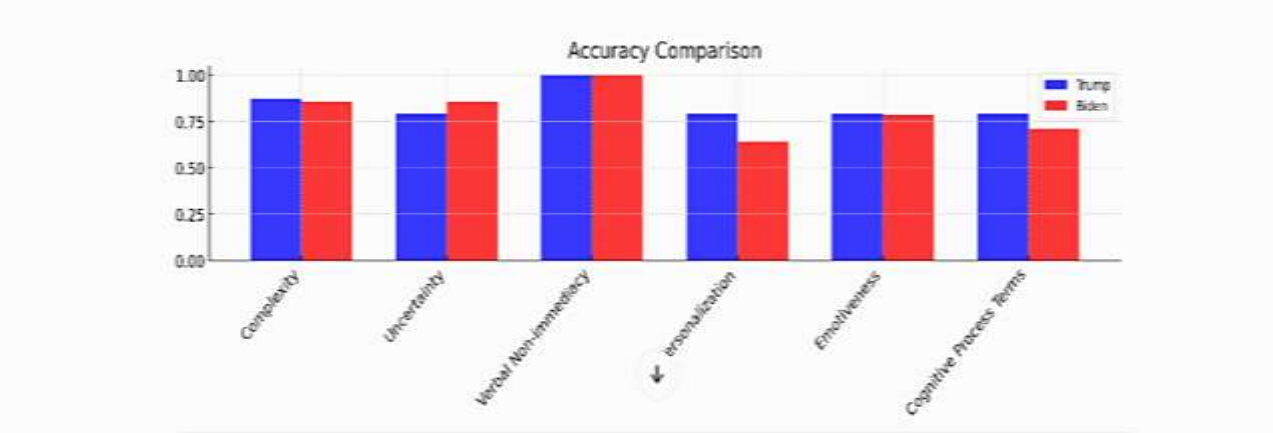

Fig. 1. Accuracy comparison across Trump's and Biden's deception markers

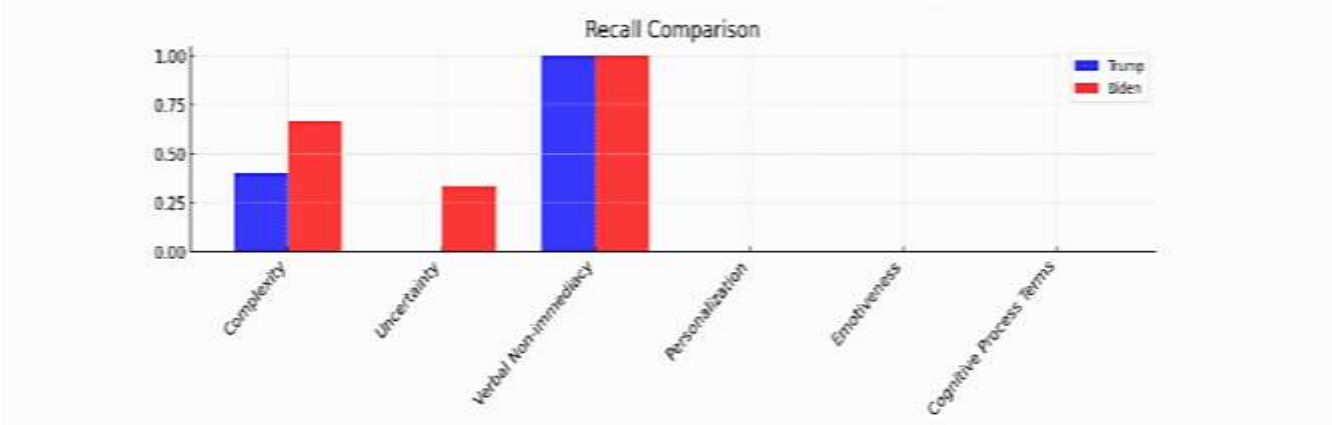

Fig. 2. Recall comparison across Trump's and Biden's deception markers

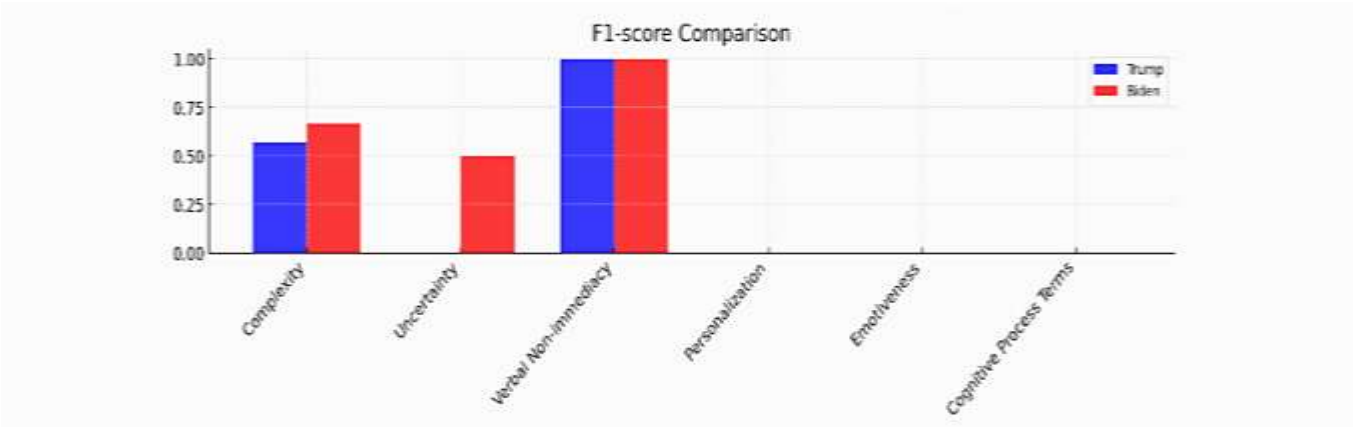

Fig. 3. F-1 score comparison across Trump's and Biden's deception markers

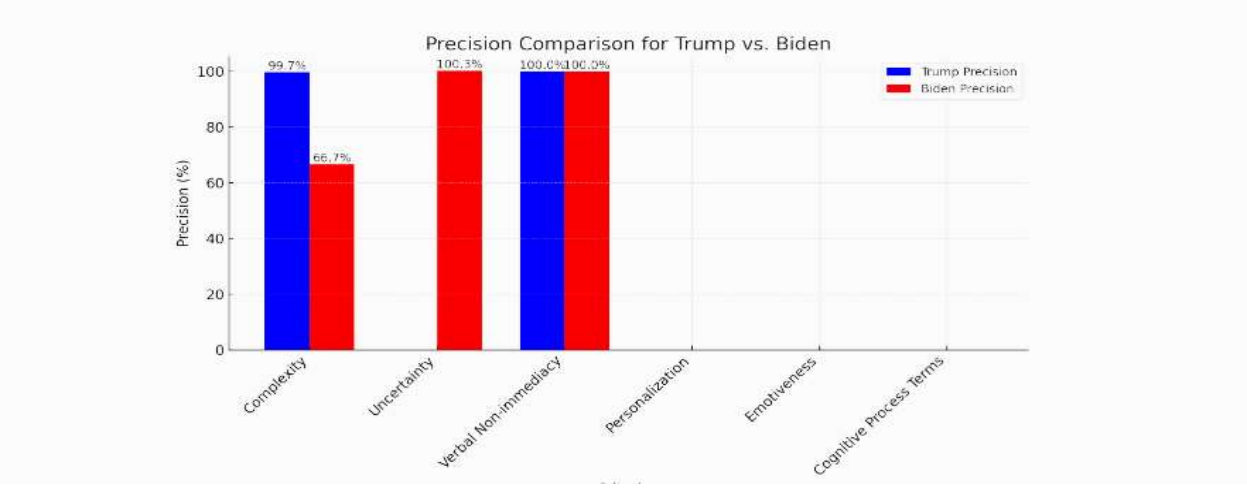

Fig. 4. Precision comparison across Trump's and Biden's deception markers

The figures 1-4 show that the accuracy comparison for both Trump and Biden exhibits high accuracy in Verbal Non-immediacy (Passive Voice), meaning that when deception is detected, passive constructions play a major role in both their statements. Moreover, complexity appears to be a slightly more accurate deception marker for Biden than for Trump, suggesting that Biden's deceptive statements may involve longer words or sentences more consistently than Trump's.

When looking at recall, Biden's deceptive statements are more frequently identified by the model, particularly in Complexity and Uncertainty. This indicates that when Biden makes deceptive statements, he is more likely to use longer words, complex sentence structures, or uncertain language (such as modal verbs and qualifiers). In contrast, Trump's recall scores are lower overall, meaning that the model struggled to identify deception in his statements, potentially due to his more direct speaking style.

The F1-score results confirm that Verbal Non-immediacy (Passive Voice) is the strongest deception indicator for both figures, reinforcing the idea that deception often involves distancing language. Biden's deceptive statements are better detected in Complexity and Uncertainty, while Trump's deceptive statements do not exhibit strong linguistic patterns in these areas. This suggests that Trump's deceptive statements might rely more on direct, categorical claims, while Biden's may involve hedging and indirectness.

6. Discussion

The findings of this study demonstrate stark differences between the Proposed Model and the Burgoon (2012) model, particularly in terms of reliability and accuracy in deception detection. The Burgoon model, achieving an accuracy of 91.84%, aligns with established research on linguistic deception detection, such as Newman et al. (2003), Hancock et al. (2008), and Fuller et al. (2011). These studies, employing regression analysis, identified deception patterns in text-based communication, with accuracy rates ranging from 70% to 80%. Despite using relatively simple statistical models, their findings have remained consistent across multiple contexts, reinforcing the idea that deceptive language follows identifiable linguistic trends. In contrast, the Proposed Model exhibited extreme inconsistencies, with an overall accuracy of 0.00%, indicating a fundamental flaw in its design. Unlike previous linguistic deception studies, which at least achieved moderate success, the Proposed Model's failure suggests severe issues with feature selection, classification methodology, or data representation.

Machine learning approaches, particularly those utilizing Support Vector Machines (SVMs), have significantly improved deception

detection accuracy by leveraging more complex feature sets. Research by Zhou et al. (2010) and Hirschberg et al. (2012) demonstrated that SVM-based models analyzing linguistic and prosodic features achieved accuracy rates between 75% and 80%, outperforming traditional regression-based methods. Additionally, multimodal SVM models incorporating speech patterns, facial expressions, and physiological signals have further improved deception detection, as seen in studies by Wang et al. (2013) and Cheng et al. (2018), which achieved classification accuracies exceeding 90%. The Burgoon model, with its balanced reliance on multiple linguistic markers and robust accuracy, aligns more closely with these high-performing machine learning models. However, the Proposed Model's failure to classify deception correctly starkly contrasts with both traditional regression models and modern machine learning approaches, suggesting that its methodology does not effectively capture deceptive linguistic features.

The discrepancies between the Proposed Model and prior research highlight key methodological concerns. Unlike successful deception detection models, which carefully select linguistic, behavioral, or multimodal features, the Proposed Model appears to overfit or misinterpret linguistic markers, leading to random or misleading results. Prior research, such as DePaulo et al. (2003) and Vrij et al. (2000), has emphasized the importance of feature reliability in deception detection. These studies found statistically significant behavioral and verbal cues through chi-square tests and logistic regression, reinforcing the necessity of feature validation before model implementation. The Proposed Model, with an accuracy of 0.00%, suggests that its feature selection lacks empirical support, further distancing it from the rigor of established deception detection methodologies.

Ultimately, while traditional statistical models (e.g., logistic regression) and modern machine learning approaches (e.g., SVMs) have consistently shown moderate to high success rates in deception detection, the Proposed Model represents a methodological failure. The Burgoon model (2012), with its high accuracy and stability, is more in line with both linguistic deception research and recent advances in AI-driven lie detection. The extreme discrepancy between the Proposed Model and prior studies suggests that more rigorous feature selection, model tuning, and multimodal data integration are necessary to improve its performance. Future research should incorporate insights from high-performing deception detection models, particularly those that effectively combine linguistic, prosodic, and physiological cues, to create a more robust and reliable deception classification system.

7. Conclusion

It can be concluded that the Burgoon model demonstrates stable and reliable performance, achieving high accuracy, precision, and recall, particularly in holistic analysis, where it records 91.84% accuracy and an F1-score of 90.91%. In contrast, the Proposed Model shows extreme inconsistencies, with an overall accuracy of 0.00%, suggesting that its predictions are dubious.

When analyzing isolated linguistic features, the Burgoon model (2012) continues to demonstrate moderate performance across multiple markers, such as Complexity, Readability, and Verb Quantity, with accuracy ranging from 50% to 70%. However, the Proposed Model produces more extreme results—while it achieves perfect classification for Verbal Non-immediacy (100% accuracy, precision, and recall), it completely fails for other key markers like Personalization, Emotiveness, and Cognitive Process Terms, which all receive 0% across precision, recall, and F1-score. This suggests that the Proposed Model may be overly reliant on a few specific linguistic patterns while failing to generalize across a broader range of deception indicators.

The analysis also reveals significant differences in deception markers between Trump’s and Biden’s statements, with Verbal Non-immediacy emerging as the strongest and most consistent predictor of deception for both figures. The high precision, recall, and F1-score for this feature suggest that both individuals tend

to rely on passive voice when making deceptive statements. Complexity also serves as a notable deception marker, with Biden's deceptive statements exhibiting greater complexity than Trump's, possibly due to longer sentence structures or increased lexical richness. In contrast, Trump's deceptive statements show a slightly higher presence of uncertainty markers, although the overall predictive power of uncertainty remains weak for both.

Despite these key patterns, the analysis highlights the limitations of some linguistic markers in deception detection. Features like Personalization, Emotiveness, and Cognitive Process Terms perform poorly, with 0% recall and precision in multiple cases, suggesting they do not reliably differentiate truthful from deceptive statements. This finding implies that deception may not always be linked to increased emotional expression or cognitive processing language, challenging common assumptions about deceptive speech. Moreover, the sharp contrast between the Burgoon (2012) model and the Proposed Model, particularly in holistic analysis, underscores the importance of model calibration and feature selection to prevent inconsistencies and improve overall reliability.

Ultimately, these findings emphasize the role of specific linguistic cues in deception detection but also highlight the need for further refinement in predictive modeling. While Verbal Non-immediacy and Complexity show promise in identifying deception, the poor performance of other markers suggests that deception is highly context-dependent and may not always manifest in predictable linguistic patterns. Future research should explore additional linguistic and contextual features to enhance model accuracy and reliability, ensuring a more comprehensive and nuanced approach to deception detection.

Declarations

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

Consent to Publish declaration: not applicable.

Consent to Participate declaration: not applicable.

Author Contribution: The paper is single-authored by Amr M. El-Zawawy. Funding Info: None.

References

1. Abouelenien, M., Pérez-Rosas, V., Mihalcea, R., & Burzo, M. (2015). Multimodal deception detection: Integrating speech, facial expressions, and physiological signals. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 589–596. https://doi.org/10.1145/2818346.2830594

2. Abouelenien, M., Pérez-Rosas, V., Mihalcea, R., Burzo, M. (2016). Deception detection using a multimodal approach. *IEEE Transactions on Information Forensics and Security, 12*(5), 1042–1055. https://doi.org/10.1109/TIFS.2016.2639337

3. Anolli, L., & Ciceri, R. (1997). The voice of deception: Vocal strategies of naive and able liars. *Journal of Nonverbal Behavior, 21*(4), 259–284. https://doi.org/10.1023/A:1024943011850

4. Arciuli, J., Mallard, D., & Villar, G. (2009). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics, 31*(3), 397–411. https://doi.org/10.1017/S014271640999017X

5. Benus, S., Enos, F., Hirschberg, J., & Shriberg, E. (2006). Pauses in deceptive speech. *Proceedings of the 3rd International Conference on Speech Prosody*, 591–594.

6. Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2012). Detecting deception through linguistic analysis. *Advances in Digital Forensics VIII*, 91–101.

7. Chen, J., Wen, Z., Xu, B., & Wang, H. (2015). A voice-based deception detection system using support vector machines. *Applied Acoustics, 89*, 99–107. https://doi.org/10.1016/j.apacoust.2014.09.005

8. Demenko, G. (2008). Voice stress analysis for deception detection. *International Journal of Speech Technology, 11*(1), 63–72. https://doi.org/10.1007/s10772-008-9031-9

9. DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

10. Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., ... & Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *Neuroimage, 28*(3), 663–668. https://doi.org/10.1016/j.neuroimage.2005.06.035

11. El-Zawawy, Amr M. (2017) Towards a New Linguistic Model for Detecting Political Lies. *Russian Journal of Linguistics*, 21 (1), 183—202.

12. Fuller, C. M., Biros, D. P., & Wilson, R. L. (2011). Decision support for deception detection: Data and text mining techniques for automated lie detection. *Decision Support Systems, 50*(3), 602–614.

https://doi.org/10.1016/j.dss.2010.08.012

13. Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes, 45*(1), 1–23. https://doi.org/10.1080/01638530701739181

14. Hirschberg, J., Benus, S., Brenier, J., Enos, F., Friedman, S., & Shriberg, E. (2012). Deceptive speech: Acoustic, prosodic, and perceptual correlates. *Proceedings of Interspeech 2012*, 1049–1052.

15. Kirchhübel, C., & Howard, D. M. (2011). Detecting deception from speech: A comparative study of three different approaches. *Speech Communication, 53*(9-10), 1076–1089.

16. Kim, J., Valstar, M. F., & Pantic, M. (2012). Fusion of facial expressions and EEG for implicit affective tagging. *IEEE Transactions on Affective Computing, 3*(2), 195–208. https://doi.org/10.1109/T-AFFC.2012.17

17. Kozel, F. A., Johnson, K. A., Grenesko, E. L., Laken, S. J., & George, M. S. (2009). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry, 68*(8), 679–686. https://doi.org/10.1016/j.biopsych.2009.05.034

18. Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping, 26*(4), 262–272. https://doi.org/10.1002/hbm.20191

19. Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

20. Pfister, T., Li, X., Zhao, G., & Pietikäinen, M. (2011). Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. *IEEE International Conference on Computer Vision (ICCV)*, 868–875.

21. Picornell, I. (2012). The linguistic markers of deception in written witness statements. *International Journal of Speech, Language & the Law, 19*(1), 1–23.

22. Rajoub, B. A., & Zwiggelaar, R. (2014). Thermal facial analysis for deception detection. *IEEE Transactions on Information Forensics and Security, 9*(6), 1015–1023. https://doi.org/10.1109/TIFS.2014.2322253

23. Reynolds, K., & Randle-Short, J. (2011). Linguistic analysis of deception in naturalistic settings. *Language and Communication, 31*(1), 51–64.

24. Vrij, A., Mann, S., Leal, S., & Fisher, R. P. (2008). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 12*(4), 141–147.

25. Vrij, A., Mann, S., & Fisher, R. P. (2000). An empirical test of the Behaviour Analysis Interview. *Law and Human Behavior, 24*(3), 313–329.

26. Wang, Y., Roesch, E. B., & Pantic, M. (2013). Automatic detection of deception in videos. *IEEE Transactions on Affective Computing, 5*(1), 58–68. https://doi.org/10.1109/T-AFFC.2013.9

27. Zhang, X., Wang, Y., Zhang, B., & Ji, Q. (2013). A multimodal approach for deception detection. *IEEE Transactions on Cybernetics, 45*(3), 492–506. https://doi.org/10.1109/TCYB.2013.2293433

28. Zhou, L., Burgoon, J. K., & Twitchell, D. P. (2004a). A longitudinal analysis of language behavior in deception. *Proceedings of the Hawaii International Conference on System Sciences*, 1–10.

29. Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology, 14*, 1–59. https://doi.org/10.1016/S0065-2601(08)60369-X

The article has been sent to the editors 14.09.25.
After processing 25.09.25.
Submitted for printing 30.09.25.