LIE DETECTION USING DEEP LEARNING

Sai Sri Harika Koundinya Vajha, Koushik Tripurari, Mohith Addepalli

1 INTRODUCTION

Lie detection refers to the process of identifying whether someone is telling the truth or not. It is a complex and challenging task, as there are many factors that can influence a person's behavior and verbal responses.

Over the years, researchers and investigators have developed various methods for detecting lies, including:

- Polygraph test: This is the most well-known lie detection technique, which measures physiological responses such as heart rate, breathing, and sweat production in response to questioning. However, its accuracy has been disputed, and it is not considered a reliable method for lie detection.
- Facial expressions: Facial expressions can provide clues about a person's emotional state, and some researchers have attempted to use them as a way of detecting lies. However, this approach has also been criticized for lacking reliability and validity.
- Eye-tracking: Some studies have suggested that eye movements can reveal whether someone is telling the truth or not. For example, liars may avoid direct eye contact or blink more frequently than truth-tellers. However, this approach is not always reliable, and some people may be able to control their eye movements when lying.
- Voice analysis: The pitch, tone, and other features of a person's voice can be analyzed to detect signs of deception. However, this method has also been criticized for its lack of reliability and validity.

2 PROBLEM DESCRIPTION

The ability to detect lies has the potential to serve a variety of important purposes, including aiding criminal investigations, ensuring national security, improving employment screening processes, and enhancing personal relationships and healthcare.

Given the potential benefits of effective lie detection, we are trying to develop solutions using deep learning neural networks, which have shown promise in achieving higher accuracy rates in this domain. These networks are capable of learning from large amounts of data, detecting patterns and relationships, and making predictions based on these insights. While there is still much work to be done in refining and validating the accuracy of lie detection using neural networks, the potential benefits of such technology make it a valuable area of research.

The focus of this research is to explore and answer the following inquiries:

1. Can we accurately predict whether a person is telling the truth or lying based on different forms of data, such as audio, video, or text?
2. To what extent does the nature of the lie being told, whether negative or positive, affect the reliability and accuracy of lie detection?
3. Are nonverbal cues, including facial expressions and body language, a dependable indicator of deception?

Through this investigation, the aim is to gain a deeper understanding of the complex and multifaceted process of lie detection and to identify potential improvements to current methods. The analysis of these questions could provide insights into the effectiveness and limitations of lie detection technologies, and inform the development of more accurate and reliable lie detection methods.

3 DESCRIPTION OF DATA:

Our dataset comprises three forms of data: video, audio, and text, which we collected from prior research conducted at Miami University.

The Multimodal Understanding of Social Interactions 3D (MU3D) is a valuable resource that consists of 321 videos. These videos feature Black and White individuals of both genders who were recorded telling the truth and lying. Specifically, there are 20 Black females, 20 Black males, 20 White females, and 20 White males, each of whom spoke honestly and dishonestly about their social relationships.

To create this dataset, each target generated four different videos, representing positive truth, negative truth, positive lie, and negative lie. This resulted in a total of 320 videos that fully encompassed target race, target gender, statement valence, and statement veracity

The video dataset is composed of 321 video files in WMV format, which include interviews conducted by professionals for lie detection purposes. These videos are labeled based on the speaker's veracity, with PT denoting Positive Truth, NT representing Negative Truth, PL indicating Positive Lie, and NL signifying Negative Lie.

The audio dataset is derived from the video files and consists of audio files stored in a separate directory.

The text dataset contains the transcripts of the video files and features a column named "transcripts." Additionally, it includes columns for valence, veracity, accuracy, and anxiousness, which measures the level of tension in the speaker's voice. This dataset serves as a valuable resource for training and evaluating machine learning algorithms designed for lie detection.

4 METHODOLOGY

Our research approach involved analyzing data from multiple modalities, namely video, audio, and text. Each modality was processed using separate CNN and RNN models, which were trained to recognize patterns in the data that corresponded to truthful or deceptive behavior.

The CNN model *(Fig 4.1: CNN Model)* were designed to identify patterns in the data by using layers such as Conv1d (1D convolution), MaxPooling (downsampling), Flatten (flatten the data), Dense (fully connected), and a final Dense layer with two possible outcomes, True (T) or Lie (L). These models were optimized to analyze data that had a spatial structure, such as video frames.

Layer (type)	Output Shape	Param #
conv1d_8 (Conv1D)	(None, 18, 32)	216128
max_pooling1d_8 (MaxPooling 1D)	(None, 9, 32)	0
flatten_6 (Flatten)	(None, 288)	0
dense_16 (Dense)	(None, 64)	18496
dense_17 (Dense)	(None, 2)	130

Total params: 234,754
Trainable params: 234,754
Non-trainable params: 0

Fig 4.1: CNN Model

On the other hand, the RNN model *(Fig 4.2: RNN Model)* was better suited for processing data with temporal structures, such as speech and text. These models used Gated Recurrent Units (GRU), which are specialized types of RNNs that can selectively retain or discard information from previous time steps. The RNN models were also designed to incorporate dropout layers to reduce overfitting and Dense layers with two potential outcomes (T or L).

Model: "sequential_8"

Layer (type)	Output Shape	Param #
conv1d_16 (Conv1D)	(None, 18, 32)	196640
max_pooling1d_16 (MaxPoolin g1D)	(None, 9, 32)	0
conv1d_17 (Conv1D)	(None, 7, 64)	6208
dropout_9 (Dropout)	(None, 7, 64)	0
max_pooling1d_17 (MaxPoolin g1D)	(None, 3, 64)	0
flatten_8 (Flatten)	(None, 192)	0
dense_16 (Dense)	(None, 128)	24704
dense_17 (Dense)	(None, 2)	258

Total params: 227,810
Trainable params: 227,810
Non-trainable params: 0

Fig 4.2: RNN Model

Once we had processed the data from each modality using CNN and RNN models, we merged the features extracted from video, audio, and text data and built a concatenated CNN model *(Fig 4.3: Concatenated CNN Model)*. This model utilized a combination of Conv1d, MaxPooling, Dropout, Flatten, Dense, and a final Dense layer with two probable outcomes (T or L). The concatenated model was trained to learn the most relevant and useful features from each modality, resulting in improved accuracy compared to the individual models.

Overall, our research approach allowed us to effectively analyze multi-modal data for the purpose of lie detection. By combining information from multiple sources, we were able to identify patterns that are difficult to detect using a single modality, leading to more accurate results. This has potential applications in a variety of domains, including law enforcement, national security, and healthcare.

```
Model: "sequential_8"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d_16 (Conv1D)          (None, 18, 32)            196640

 max_pooling1d_16 (MaxPoolin  (None, 9, 32)            0
 g1D)

 conv1d_17 (Conv1D)          (None, 7, 64)             6208

 dropout_9 (Dropout)         (None, 7, 64)             0

 max_pooling1d_17 (MaxPoolin  (None, 3, 64)            0
 g1D)

 flatten_8 (Flatten)         (None, 192)               0

 dense_16 (Dense)            (None, 128)               24704

 dense_17 (Dense)            (None, 2)                 258

=================================================================
Total params: 227,810
Trainable params: 227,810
Non-trainable params: 0
_________________________________________________________________
```

Fig 4.3: Concatenated CNN Model

5 RESULTS:

Upon completion of the project, we were able to successfully answer the three research questions that were posed:

Can we accurately predict whether a person is telling the truth or lying based on different forms of data, such as audio, video, or text?:

To what extent does the nature of the lie being told, whether negative or positive, affect the reliability and accuracy of lie detection?

Are nonverbal cues, including facial expressions and body language, a dependable indicator of deception?

Data variety	CNN (Training and Test Acc)	RNN(Training and Test Acc)
Video+Audio+Text	Train Accuracy:75% Test Accuracy: 56%	Train Accuracy:55% Test Accuracy: 51%
Video+Audio	Train Accuracy:75% Test Accuracy: 53%	
Video	Training Accuracy: 65% Testing Accuracy: 51%	
Audio	Training Accuracy: 39% Testing Accuracy: 30%	
Text	Training Accuracy: 60% Testing Accuracy: 50%	Training Accuracy: 56% Testing Accuracy: 43%

Fig 5.1: Results achieved for models

Regarding the first question, we found that it is indeed possible to accurately predict whether an individual is lying or telling the truth based on various forms of data, including audio, video, and text. However, we discovered that the accuracy of the model depends on how the individual features are analyzed. When the features were taken into account individually, the model's performance was good. However, the best results were obtained when the features were concatenated.

The second research question focused on how the nature of the lie, whether positive or negative, affects the reliability and accuracy of lie detection. Our findings indicated that when we considered only two labels, True (T) or Lie (L), the accuracy of the model was much higher than when we considered four labels, Positive Truth, Positive Lie, Negative Truth, and Negative Lie. As part of the future scope of the project, we plan to explore methods to improve accuracy when considering all four labels.

Lastly, we investigated the reliability of nonverbal cues, including facial expressions and body language, as indicators of deception. Our research found that these nonverbal cues could indeed be dependable indicators of deception. Specifically, we found that the facial expressions and body language of individuals could be effectively used to identify deception through a video processing CNN model.

REFERENCES:

E. Paige Lloyd & Jason C. Deska & Kurt Hugenberg & Allen R. McConnell & Brandon T. Humphrey4 & Jonathan W. Kunstman : Miami University deception detection database

Albrechtsen, J. S., Meissner, C. A., & Susa, K. J. (2009). Can intuition improve deception detection performance?. Journal of Experimental Social Psychology, 45, 1052–1055.

Levine, T. R. (2007). NSF funded cheating tape interviews. East Lansing, MI: Michigan State University

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. Personality and Social Psychology Review, 10, 214–234.

Gray, H. M., Mendes, W. B., & Denny-Brown, C. (2008). An in-group advantage in detecting intergroup anxiety. Psychological Science, 19, 1233–1237.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? Perspectives on Psychological Science, 7, 528–530.

Pauker, K., & Ambady, N. (2009). Multiracial faces: How categorization affects memory at the boundaries of race. Journal of Social Issues, 65, 69–86.

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. Social Cognition, 24, 607–640.

Shoda, T. M., & McConnell, A. R. (2013). Interpersonal sensitivity and self-knowledge: Those chronic for trustworthiness are more accurate at detecting it in others. Journal of Experimental Social Psychology, 49, 440–443.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. Applied Cognitive Psychology, 11, 373–397.

ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some evidence for unconscious lie detection. Psychological Science, 25, 1098–1105. https://doi.org/10.1177/0956797614524421

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. Journal of Nonverbal Behavior, 24, 239–263.

APPENDIX 1: CONTRIBUTION OF EACH MEMBER

Koushik Tripurari came up with the idea to develop models for lie detection. He was responsible for conceptualizing and developing the CNN model for analyzing video and audio data. Convolutional Neural Networks are a type of deep learning model that are often used for image and audio analysis. Koushik's contribution was crucial to developing a comprehensive lie detection model that could analyze a variety of data modalities.

Sai Sri Harika Koundinya Vajha made significant contributions to the field of deception detection by building upon CNN models for video and concatenated models. Her primary focus was on enhancing the accuracy of these models, which she achieved by fine-tuning the CNN models to make them more effective at detecting deception. As a result, she was able to present her research findings, which included the improved performance of these models. Overall, her work has contributed to advancing the field of deception detection and has the potential to be beneficial in various domains where lie detection is crucial.

Mohith Addepalli played a key role in obtaining the necessary data for the project. He also focused on developing models for analyzing text data, which is another important modality for lie detection. Natural Language Processing (NLP) models can be used to analyze written or spoken language and identify patterns of deception.

Together, the team worked collaboratively to develop and refine lie detection models using various data modalities. By combining their unique skills and expertise, they were able to create a more comprehensive and accurate approach to detecting lies.