Deception Detection using Machine Learning and Computer Vision Approaches

University of Peradeniya

T.Sankavi [1]

Supervised by: Prof. Amalka Pinidiyaarachchi[1]

[1]*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya*

Abstract - The study develops an automated system for detecting deceptive behavior using visual and vocal cues from the Miami University Deception Detection dataset, which includes 320 labeled video samples. It employs deep learning for feature extraction and various machine learning algorithms for classification, achieving around 70% accuracy with the Random Forest algorithm. Eye movements are highlighted as significant indicators of deceptive behavior.

Introduction

Deception is a complex human behavior that has a significant impact in various domains, including law enforcement, intelligence operations, security, psychology, business negotiations, and personal relationships. The ability to correctly distinguish between honest and deceptive communication is important to identify fraudulent activities and maintain trust and authenticity. However, deception detection remains a complex and challenging task, as individuals vary in their natural behaviors, making it challenging to establish universal cues for deception and skilled deceivers can control their verbal and nonverbal cues, making it hard to identify inconsistencies or signs of lying. Thus it is challenging for humans to determine whether a person is being deceptive. Traditional approaches to deception detection, such as intuition and behavioral cues, have proven to be unreliable and prone to biases. Therefore, throughout history, many methods and devices were developed for that task.

Objectives

- Develop a Machine Learning model capable of detecting deceptive behavior based on visual and linguistic features.
- Investigate facial micro-expressions to identify the most prominent area of the face for deceptive behavior

Methodology

Machine Learning Model

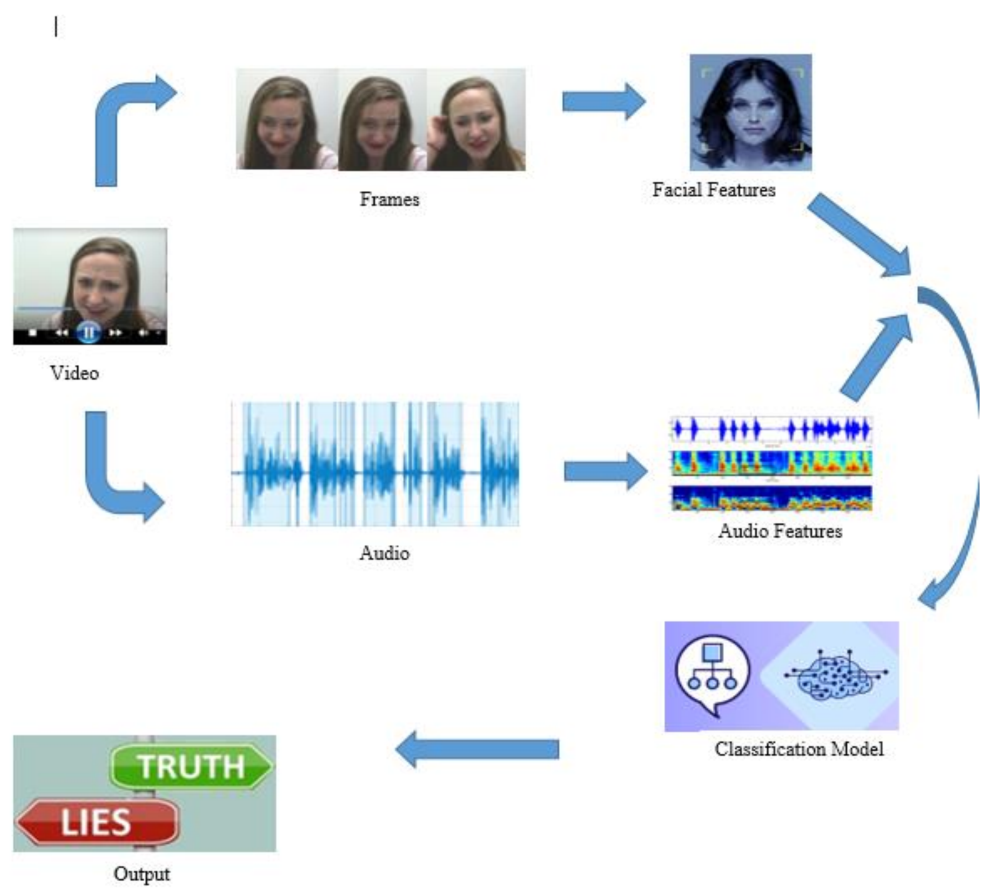


Figure 1: Machine Learning Model

The initial step involves extracting frames and audio from the source, capturing both visual and vocal components. Following this, preprocessing techniques are applied to standardize and refine the data. Subsequently, the process includes extracting audio features and visual features to glean pertinent information from the respective signals. The extracted features serve as inputs for training a classification model, employing Support Vector Machine, Random Forest, and Artificial Neural Networks. The model is trained on the dataset, and its performance is evaluated.

Transfer Learning Model

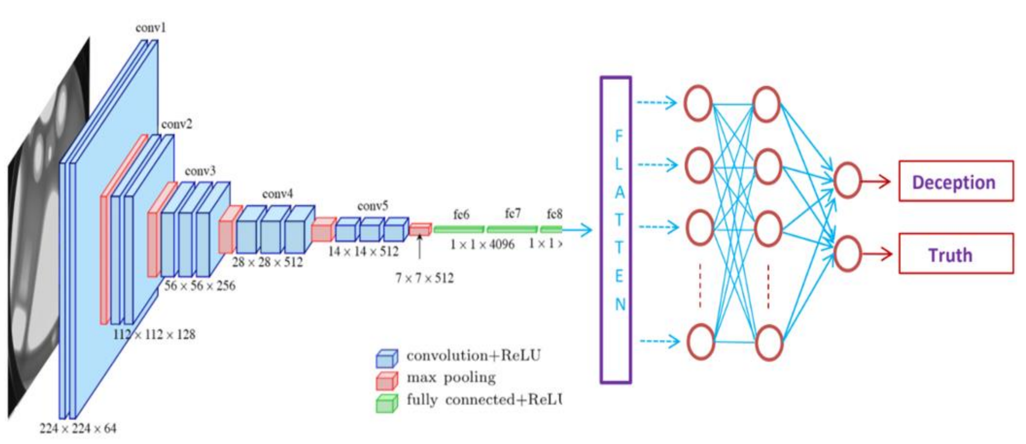


Figure 2: Transfer Learning Model

	Layer	Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	224x224x3	-	-	-
1	2 x Convolution	64	224x224x64	3x3	1	ReLU
	Max Pooling	64	112x112x64	3x3	2	ReLU
3	2 x Convolution	128	112x112x128	3x3	1	ReLU
	Max Pooling	128	56x56x128	3x3	2	ReLU
5	2 x Convolution	256	56x56x256	3x3	1	ReLU
	Max Pooling	256	28x28x256	3x3	2	ReLU
7	3 x Convolution	512	28x28x512	3x3	1	ReLU
	Max Pooling	512	14x14x512	3x3	2	ReLU
10	3 x Convolution	512	14x14x512	3x3	1	ReLU
	Max Pooling	512	7x7x512	3x3	2	ReLU
13	FC	-	25088	-	-	ReLU
14	FC	-	512	-	-	ReLU
15	FC	-	512	-	-	ReLU
Output	FC	-	2	-	-	Softmax

Architecture of the Transfer Learning Model

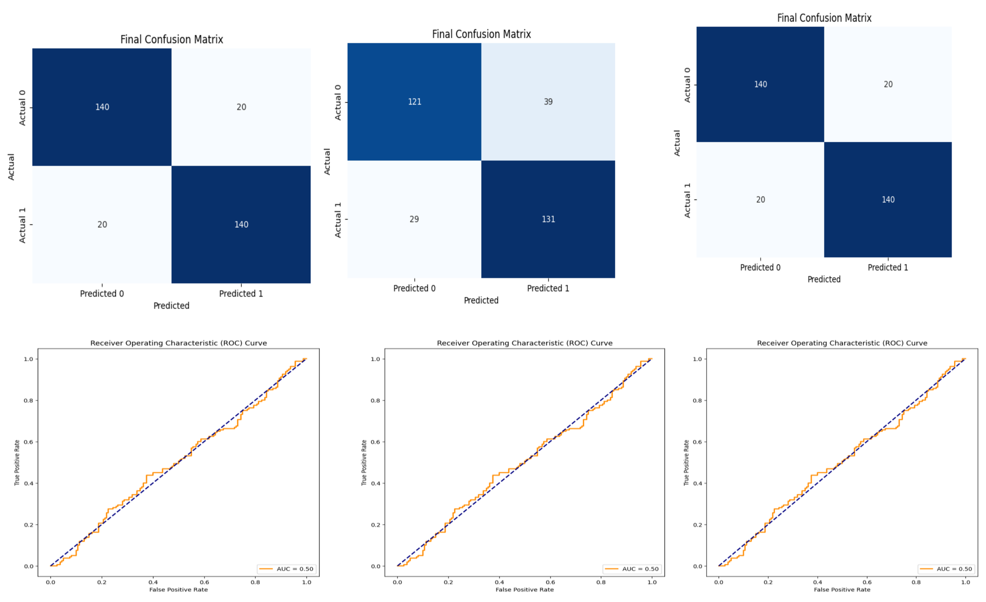
The model employs a pre-trained VGG16 convolutional neural network, as a feature extractor. Extracted frames are resized to 224x224. After splitting the data into training and testing sets, label encoding and one-hot encoding prepare the labels. The VGG16 base model, excluding classification layers, forms the foundation for a new Sequential model. This model includes layers like Flatten, Dense with ReLU activation, Dropout for regularization, and a final Dense layer with softmax activation for binary classification. Compiled using Adam optimizer and categorical cross-entropy loss.

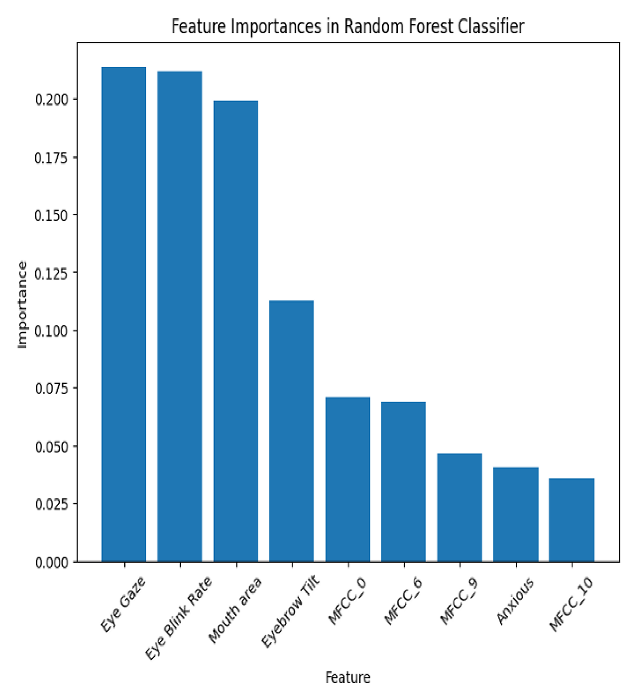
Results and Discussion

Machine Learning Model

Train/Test	Acc(SVM)	Acc(RF)	Acc(ANN)
R1	0.71	0.74	0.66
R2	0.71	0.70	0.63
R3	0.71	0.70	0.66
R4	0.71	0.74	0.65
R5	0.78	0.77	0.74
R6	0.65	0.74	0.61
R7	0.61	0.77	0.60
R8	0.61	0.70	0.63
R9	0.68	0.74	0.70
R10	0.77	0.70	0.687
Avg.Acc	**0,70**	**0.74**	**0.69**

Accuracy in each iteration of the 10-fold cross-validation approach, along with average accuracy across the 10 runs of the 10 runs





Train/Test runs	Accuracy (SVM)	Accuracy (RF)	Accuracy (ANN)
R1	0.71	0.74	0.61
R2	0.65	0.80	0.61
R3	0.70	0.80	0.67
R4	0.75	0.78	0.38
R5	0.70	0.80	0.50
R6	0.71	0.74	0.57
R7	0.65	0.70	0.60
R8	0.68	0.75	0.68
R9	0.71	0.72	0.60
R10	0.53	0.65	0.63
Avg.Acc	**0.70**	**0.72**	**0.59**

Results from Recursive Train/Test Runs for visual features only

Train/Test runs	Accuracy (SVM)	Accuracy (RF)	Accuracy (ANN)
R1	0.71	0.68	0.61
R2	0.59	0.50	0.61
R3	0.56	0.56	0.80
R4	0.59	0.65	0.46
R5	0.53	0.43	0.57
R6	0.68	0.62	0.61
R7	0.65	0.65	0.60
R8	0.68	0.50	0.63
R9	0.71	0.56	0.60
R10	0.53	0.65	0.56
Avg.Acc	**0.63**	**0.58**	**0.61**

Results from Recursive Train/Test Runs for vocal features only

Transfer Learning Model

Epoch	Accuracy	Loss	Val_acc	Val_loss
1	0.0788	0.4987	0.6938	0.4946
2	0.6965	0.4991	0.6931	0.5054
3	0.6933	0.4991	0.6931	0.5054
4	0.6932	0.5031	0.6931	0.5054
5	0.6932	0.5031	0.6931	0.5054
6	0.0765	0.4890	0.6890	0.4923
7	0.6965	0.4991	0.6631	0.5318
8	0.6823	0.4991	0.6789	0.5054
9	0.6932	0.4928	0.6931	0.5290
10	0.6943	0.4932	0.6765	0.5356

Loss and Accuracy during training and validation for each epoch

The training loss decreases significantly over epochs. Both training and validation accuracies are around 50%, indicating limited ability to distinguish between deceptive and non-deceptive instances. The test accuracy, at 52.66%, aligns with the training and validation performance, indicating that the model's struggles are consistent across different data.

Conclusion

The results of the deception classification based on both visual and vocal features present a comprehensive evaluation of the performance of three different machine learning models: Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). The overall accuracy rates for SVM, RF, and ANN are reported as 0.70, 0.74, and 0.69, respectively, providing an initial overview of their effectiveness in discerning between deceptive and non-deceptive behaviors.

Contact details

Name: Dr. Erunika O. Dayartna

Tel. No.: +94 76 049 5054

Email : erunika.dayaratna@sci.pdn.ac.lk

Multidisciplinary AI Research Centre (MARC)

University Research Council

University of Peradeniya

Peradeniya, 20400, Sri Lanka