Deception Detection using Machine Learning

DDP Project Report
Submitted in partial fulfillment of the requirements
for the degree of

Bachelor of Technology
(Mechanical Engineering)
and
Master of Technology
(Computer Integrated Manufacturing – CIM)

By
Pattisapu Divya
(Roll No. – 160040084)

Under the guidance of
Prof. Asim Tewari

Department of Mechanical Engineering
Indian Institute of Technology Bombay
Mumbai

Approval Sheet

This report entitled "**Deception Detection Using Machine Learning"** by Pattisapu Divya with Roll No. 160040084 is approved for the degree of Dual Degree (B.Tech + M. Tech).

Examiners

Prof. Makarand Kulkarni ________________________

Prof. Anirban Guha Anirban Guha

Supervisor

Prof. Asim Tewari Asim Tewari

Chairman

Prof. Makarand Kulkarni________________________

Date: July 9, 2021

Place: Mumbai

Acceptance Certificate

It is hereby certified that this Dual Degree Project Stage-2 Report titled "Deception Detection using Machine Learning" by Pattisapu Divya is approved by me for submission. Certified further that, to the best of my knowledge, the report represents work carried out by the student.

Prof. Asim Tewari

Date:

Declaration of Academic Integrity

"I declare that this written dissertation represents my ideas in my own words and where others' copyright ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my dissertation. I understand that any violation of the above will be cause for disciplinary action as per the rules of regulations of the Institute"

Date: Signature

Pattisapu Divya

(160040084)

Acknowledgement

I would like to express my sincere gratitude towards my guide Prof. Asim Tewari for his invaluable support and guidance throughout my Stage 1 and Stage 2. Most importantly, he has given me a freehand in this project, thereby giving me the freedom to explore and pursue various directions to address the problem at hand. I would also like to thank RA – Shefali Gokarna. Without her help and support, this report would not have taken the present shape.

Pattisapu Divya

(160040084)

Table of Contents

List of Figures

List of Tables

Abstract

This dissertation describes work on deception detection in text speech and video using machine learning techniques. Deception detection has been a study of interest in various domains like forensics, psychology, computer science and linguistics. It has promising applications in real life as the human counterparts are poor at detecting lies. This work examines the ability to deceive for different types of questions: Leading questions and open-ended questions, positive assumptions and negative assumption questions, outcome and process questions, recollection and hypothetical questions. This dissertation shows that video processing techniques and speech processing techniques can be applied to the domain of deception to extract useful features for classification. We also looked at the personality factors using the big five personality test and checked its impact on deception detection ability. An important product of this work is a balanced video and audio data set simulating a job interview. Results show that supervised classification models like K nearest neighbours classifier and support vector classifier can classify text and audio data to some success: The accuracy is better than human hearers and chance. Long short-term machines are used to classify the video data after facial feature extraction leading to test accuracy of 64.7%.

1. Introduction

1.1.Deception

Deception is defined as a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver. [4] Deception occurs when communicators control the information contained in their messages so that they convey a meaning that is different from the truth that is known to them. This definition excludes mistaken or unintended lies. [5] classifies deception into the following categories.

Type of lie	**Explanation**
Avoidance	attempts to escape or minimize negative consequences associated with specific behaviors
Concealment	involve ongoing deceptions in which people misrepresent qualities within themselves
Gainful-Falsification	written lies employed to extract a benefit from another person or institution
Gainful-Malice	a class of Verbal-Malice deceptions in which the extraction of benefit becomes the key element of the deception
Gainful-Misleading	lies that are employed to extract a specific benefit from another person
Interpersonal-Ploy	deceptions employed within the space of an ongoing interaction to improve the pleasantness of that interaction
Social-Enhancement	deceptions employed to improve one’s social standing by impressing or gaining sympathy from others
Verbal-Trickery	told for self-serving purposes, and typically involve a specific harm to another person
Verbal-Malice	lies in which one endorses a particular course of behavior but then proceeds to engage in a less socially desirable behavior

Table 1. 1 Categories of deception

1.2.Deception in an interview setting

Deception is a common phenomenon in everyday life. On average studies show that people report line one or two times per day [3]. [2] self-presentation approach suggests that people purposely regulate their verbal and nonverbal behaviors for impression management. Deceivers and truth-tellers are equally concerned with impression management because people want to appear likeable, attractive, or interesting, more so in an interview setting where first impressions matter. In an interview setting, according to [6], applicants lie both on job applications and during the interview primarily to appear to conform to job requirement. Avoidance, Concealment and Gainful-Falsification (Table 1.1) are common interview deception categories.

1.3.Deception detection - Motivation

People are generally poor lie detectors. A meta-analysis by [1] shows that, on average, subjects in 206 studies perform near chance. So, an automatic detection deception system is capable of performing better than an average human regardless of training. Thus, deception detection is a unique problem where the gold standard is not human performance. It has received attention from diverse fields like psychology, philosophy, sociology, criminology, linguistics, computer science and has found application in forensics, law enforcement, military, and intelligence Agencies. Airports, courts, police interrogations rely on decisions concerning deceptive behaviour but are currently subject to human errors and bias. Traditional physiological methods like polygraphs have failed in multiple cases which has resulted in falsely accusing the innocent, or letting the guilty roam free. Invasive techniques such as MRI or similar invasive sensors are not feasible due to inconvenience of handling equipment and high cost. Thus, automatic deception detection using behavioral cues has immediate applications in real life. This domain is relatively new and interesting and it has a large scope for improvement.

1.4.Idea

The long term goal is the deployment of real time deception detection model which can take video and audio input and immediately predict the veracity of the statement made. This can be useful for online interviews courtroom sessions police interrogation Security identity verification in immigration offices. Moreover, analysis of the model can tell which features give away a deception. This can be helpful for training detectives or those working infields that require deception detection. With the advent of technology more people communicate across the world using computer mediated communication (CMC). A text-based model can fact check online reviews, social media posts and online interview applications. It might also help one become a good liar if they gain control of these cues, useful for undercover operations.

1.5.Goal

The work aims to understand behavioral changes when a person tries to deceive or lie. We want to explore various factors: Verbal, nonverbal, outcome, personality and demographic, which influence the ability to deceive in an individual. We want to understand what kind of questions trigger lies that are easy to detect. Finally, we want to create an automatic deception detection (ADD) model using the verbal and behavioral input of the person of interest, which can classify speech, text and video into truth and lie.

1.6.Outline

Chapter 1 (current chapter) gives a basic introduction of the project and explains the common terms mentioned in the topic, along with the motivation to pursue it.

In Chapter 2, Relevant datasets are tabulated in Section 2.1. Literature Review is done separately for text analysis literature in Section 2.2, Audio analysis literature in Section 2.3, Video analysis literature in Section 2.4, and Personality factors literature in Section 2.5.

In Chapter 3, the problem statement is chalked out in brief detail. The corpus is described in detail, along with the approach taken to analyse it.

In Chapter 4, modeling and implementation of our work are explained in high detail for more lucid understanding. It is explicitly divided into three main subsections: text, audio and video along with final results.

In Chapter 5, Conclusions and possible future directions of this work are presented.

2. Literature Review

2.1.Datasets

Sr. No.	Dataset Name	Paper Name	Type	Number of people	Total number of questions answered	Total size (if applicable)	References
1	CXD Corpus	Deep personality detection for deception detection	Audio	346	8304	125 hours	[7]
2	Casual Deception Data	Bag of lies - A multimodal dataset for deception detection	Audio, Video	35	325	-	[8]
3	CSC Corpus	Combining Prosodic Lexical and Cepstral Systems for Deceptive Speech Detection	Audio	32	6 areas	-	[9]
4	Lie Detection	LieToMe: An ensemble approach for deception detection from facial cues	Video	66	121	57 minutes	[10]
5	**CDC Corpus**	**Cross culture production and detection of deception from speech**	**Audio**	**278**	**6672**	100.5 hours	[11]
6	Employee Selection	Text analysis software to detect deception in written short answer questions	Text	106	106	32000 words	[12]
7	Real life trial dataset	Detecting deception language in crime interrogation	Text	496	-	-	[13]

8	Impression Management Strategy in Job Interviews	Looking good and lying to do it	-	59	-	-	[6]
9	Items Packed in Suitcase	Deception in the eyes of deceiver - a CV and ML based automated deception detection	Video	100	1200	255000 vectors	[14]

Table 2. 1 Datasets

2.2.Text analysis literature

The literature on verbal cues to deception indicates that narrative which is fabricated may differ from narrative which is truthful at every level from global discourse to individual word choice. Features of (a) narrative structure and length, (b) text coherence, (c) factual and sensory detail, (d) filled pauses, (e) syntactic structure choice, (f) verbal immediacy, (g) negative expressions, (h) tentative constructions, (i) referential expressions, and (j) particular phrasings have all been shown to differentiate truthful from deceptive statements in text [15]

2.2.1.Understanding Deception Text Features

[16] explains the following useful features in detail, while analysing the texts from "real world" sources—criminal statements, police interrogations and legal testimony. This paper was manually tagged for 12 linguistic indicators of deception that were cited in the psychological and criminal justice literature by a team of linguistics.

Reason	Categories	Examples
Lack of commitment to a statement or declaration. Purpose: to avoid making a direct statement of fact.	linguistic hedges including non-factive verbs and nominals; describes words whose meaning involves	I guess so, maybe, and sort of. my understanding and my recollection, possible and approximately,

	ambiguity,	something a glimpse and between 9 and 9:30.
	qualified assertions, which leave open whether an act was performed	I needed to finish my thesis;
	unexplained lapses of time	later that day;
	overzealous expressions	I swear to God
	rationalization of an action	I was unfamiliar with the road
Preference for negative expressions	negative forms, either complete words such as never or negative morphemes as in inconceivable;	Not really, never
	negative emotions	I was a nervous wreck
	memory loss	I forget.
Inconsistencies with respect to verb and noun forms. Four of the indicators make up this class:	verb tense changes	I can't do enough. My children wanted me. They needed me. And now I can't help them.
	thematic role changes	They might change the thematic role of a Noun Phrase from agent in one sentence to patient in the following sentence
	noun phrase changes	Different Noun Phrase forms are used for the same referent or to change the focus of a narrative;
	pronoun changes	similar to noun phrase changes changes in tense are often more indicative of deception than the overall choice of tense.

Table 2. 2 Text deception features

2.2.2.Using LIWC software for text feature extraction of various categories

[12] asks 106 Participants to answer two short-answer questions both deceptively and honestly in online recruitment settings. Two questions were asked about a personal achievement and ability to collaborate, respectively:

Model	Hypothesis	Reason	associated LIWC

			parameter
The Newman-Pennebaker (NP) model	fewer first person singular pronouns	due to psychological distancing	pronoun
	fewer conjunctions	a measure of language complexity	conjunc
	more motion verbs	descriptors with little cognitive load	motion
	more negative emotion words	leakage hypothesis (Ekman, 2007; Hancock & Woodworth, 2013), the person lying is experiencing emotions such as guilt or shame, due to lying are which cause the person to exhibit non-verbal leakage that goes against the words/expressions that he is consciously trying to control	negativ
The Reality Monitoring (RM) framework	more sensory, spatial, temporal, and affective words	theorized to reflect the qualities of a true memory	
	less Cognitive process words	lying is cognitively taxing	affect, time, see, hear, space

Table 2. 3 Text deception models and corresponding LIWC features

Predictor	β	SE	Wald	p	Odds ratio	95% CI for odds ratio Lower	Upper
Q1 Personal achievement							
First person Singular pronoun	−.07	.04	2.23	.14	.94	.86	1.02
Negative emotions	−.30	.15	4.09	.04	.74	.55	.99
Conjunctions	−.16	.08	3.66	.06	.85	.73	1.00
Motion verbs	.16	.11	2.20	.14	1.17	.95	1.45
Intercept	1.50	.66	5.08	.02	4.40		
Q2 Collaboration							
First person Singular pronoun	−.02	.05	1.00	.75	.98	.75	.98
Negative emotions	.33	.15	4.41	.04	1.39	.04	1.39
Conjunctions	−.12	.08	2.49	.11	.88	.11	.88
Motion verbs	.24	.12	3.85	.05	1.27	.05	1.27
Intercept	.18	.60	.09	.76	1.19		
Q1 & Q2 Combined responses							
First person Singular pronoun	−.07	.06	1.12	.29	.93	.82	1.06
Negative emotions	−.02	.18	.01	.91	.98	.69	1.40
Conjunctions	−.20	.09	4.20	.04	.82	.68	.99
Motion verbs	.33	.15	4.80	.03	1.40	1.03	1.87
Intercept	1.12	.77	2.12	.15	3.06		

Note: NP model consists of the first person singular pronouns, conjunctions, motion, and negative emotions.

Figure 2. 1 Logistic regression coefficients for Newman-Pennebaker model

The logistic regression coefficients of the NP model are shown in Figure. From the coefficients, conjunctions and motion verbs were both significant and in the hypothesized direction, thus indicating that deceptive language contained fewer conjunctions and a more use of motion verbs. CohMetrix is suggested as a promising software program that has a better sophistication than LIWC.

	β	SE	Wald	p	Odds Ratio	95% CI for odds ratio Lower	Upper
Q1 Personal achievement							
Cognitive processes	−.06	.05	1.70	.19	.94	.86	1.03
Affect	−.04	.07	.39	.53	.96	.84	1.09
Time	−.08	.06	1.90	.17	.92	.82	1.04
Space	−.02	.07	.14	.71	.98	.86	1.11
See	−.11	.21	.28	.59	.89	.60	1.34
Hear	.36	.26	1.88	.17	1.43	.86	2.40
Intercept	1.44	1.01	2.04	.15	4.24		
Q2 Collaboration							
Cognitive processes	−.05	.04	1.50	.22	.95	.88	1.03
Affect	.24	.09	7.77	.01	1.27	1.07	1.51
Time	−.01	.07	.04	.83	.99	.86	1.13
Space	.09	.07	1.52	.22	1.09	.95	1.25
See	.14	.28	.27	.60	1.16	.67	2
Hear	.63	.34	3.36	.07	1.87	.96	3.67
Intercept	−1.31	1.03	1.63	.20	.27		
Q1 & Q2 Combined responses							
Cognitive processes	−.09	.05	2.95	.09	.91	.82	1.01
Affect	.12	.10	1.35	.24	1.12	.92	1.37
Time	−.12	.09	1.90	.17	.89	.75	1.05
Space	.04	.09	.17	.68	1.04	.87	1.24
See	.05	.33	.02	.89	1.05	.55	2
Hear	1.03	.43	5.60	.02	2.80	1.19	6.54
Intercept	.52	1.33	.15	.69	1.69		

Note: RM model consists of sensory, spatial, temporal, affective, and cognitive processes.

TABLE 3 Logistic regression of reality monitoring model predicting deception

Figure 2. 2 Logistic regression coefficients for the Reality Monitoring Model

2.2.3.Using NLP (Natural Language Processing) text features

[17] analyzes the CXD corpus using several feature extraction techniques.

1. Dictionary of Affective Language Features: DAL is a lexical analysis dictionary that is used for investigating emotive content of speech. It lists around 4500 English words, along with a rating for Pleasantness and for Activation (Arousal) associated with every word.

2. The complexity of an utterance is calculated as the number of syllables per speech segment divided by the number of words
3. Bool features capturing whether the utterance contains a feeling word, hedge word, number, or date
4. Contains contraction, or a yes/no answer instead of complete sentence.
5. a bag of words model is made using part of speech tags obtained using NLTK's built-in POS tagger.

Features	UAR	+baseline
DAL	63.1	61.5
LIWC	63.9	61.9
FFV	54.3	60.7
Phonotactic	**67.7**	61.9
POS	57.8	57.8
Lexical	59.3	61.3
Duration	56.9	**62.2**
Baseline (majority)	50	-
Baseline (openSMILE)	61.9	-

Figure 2. 3 Results for CXD corpus modal analysis

Binary features like isDate, isYes perform well in the model.

2.2.4.Using LSTMs and MLP

[7] On the CDC corpus, a multilayer perceptron (MLP) is trained using LIWC, DAL, LLD, and pretrained word embeddings, a Long Short-Term Memory classifier using Stanford's pretrained GloVe vectors, and a hybrid of these two models. After training the word embedding layer using initialization from GloVe pre-trained model, we feed 300-dimensional word embeddings one at a time to the LSTM layer to get instance-level representations. A sigmoid function is then applied outputting a probability estimation of the instance's deception or lack thereof.

Table 2: *Deception detection without personality*

Model	Precision	Recall	F1
MLP	68.08	67.95	68.01
LSTM	65.64	66.08	65.78
Hybrid	69.43	69.46	69.45

Table 3: *Multi-task learning: variant (i)*

Model	Precision	Recall	F1
MLP	74.33	74.51	74.39
LSTM	64.61	65.56	64.40
Hybrid	69.42	69.67	69.51

Table 4: *Multi-task learning: variant (ii)*

Model	Precision	Recall	F1
MLP	74.37	74.67	74.38
LSTM	66.13	67.03	65.89
Hybrid	72.58	72.98	72.70

Table 5: *Deception detection using gold-standard personality as a late feature*

Model	Precision	Recall	F1
MLP	70.08	70.00	70.04
LSTM	65.09	65.74	65.22

Figure 2. 4 Results for MLP and LSTM text feature models

2.3. Audio analysis literature

2.3.1. Using Acoustic-Prosodic features

[11] analyzes acoustic and prosodic speech features on the CDC corpus. The unit of analysis is an inter-pausal unit (IPU), defined as a pause-free segment of speech from a speaker. 50 ms of silence is considered as a threshold for pause. IPUs are created using the software Praat. Personality (NEO-FFI) features and gender features were included as part of the model.

Acoustic-Prosodic Features	**Physical significance**
f0 minimum, f0 maximum, f0 mean, f0 median, f0 standard deviation, f0 mean absolute slope	the physical correlate of pitch
intensity minimum, intensity maximum, intensity mean, intensity standard deviation	measures of a correlate of perceived loudness
jitter, shimmer, noise to harmonics ratio	measures of voice quality, variation in vocal fold behavior which might lead to listeners' perception as the harshness or creakiness or breathiness of the voice

Table 2. 4 Acoustic-Prosodic features and physical significance

Table 1. Accuracy of 3 models, using raw acoustic-prosodic features and 2 methods of feature normalization

Model	Raw	SessionNorm	BaselineNorm
J48	59.89	62.09	62.19
Bagging	58.65	61.19	61.01
RandomForest	61.23	**63.03**	62.79

Figure 2. 5 Results for different normalization methods: Audio data

Session normalization or normalization per person and baseline normalization, using 3-4min of normal speech from the person are the two normalization techniques used.

Table 2. Results using session normalized features and personality scores, gender and language

Model	SessionNorm + NEO, gender, lang
J48	64.86
Bagging	63.9
RandomForest	**65.86**

Figure 2. 6 Results for audio, text, gender and personality data

2.3.2.Using Phonotactic features

[17]This feature deals with restrictions in a language on the permissible combinations of phonemes. It defines a permissible syllable structure, consonant clusters and vowel sequences. Deceptive speech may result in differences in pronunciation between individuals,i.e. a deceptive speaker may tend to choose certain phonotactic variants or words more frequently than others. An English phoneme recognizer developed by Brno University of Technology (BUT) is applied to generate the phoneme hypotheses for an audio instance. A trigram language model is built with Witten-Bell smoothing for each class in training set, using SRILM. WordLM and phonLM are two models that extract phonotactic features. It

performs well on validation set, but isn’t easy to generalize since it is speaker dependent.

Table 1: *Deception classification results on dev set, using single feature sets and single+baseline feature sets*

Features	UAR	+baseline
DAL	63.1	61.5
LIWC	63.9	61.9
FFV	54.3	60.7
Phonotactic	**67.7**	61.9
POS	57.8	57.8
Lexical	59.3	61.3
Duration	56.9	**62.2**
Baseline (majority)	50	-
Baseline (openSMILE)	61.9	-

Figure 2. 7 Phonotactic feature model results for CDC corpus

2.3.3.Using BLSTMs.

[18] analyzes the CXD corpus, a collection of within-subject deceptive and non-deceptive speech from native speakers of Standard American English and Mandarin Chinese. They use the Interspeech 2013 ComParE Challenge base-line feature set, which contains 6373 features from the computation of various functionals over low-level descriptor contours extracted from openSMILE python library. The lexical features include n-grams and GloVe pre-trained embeddings.

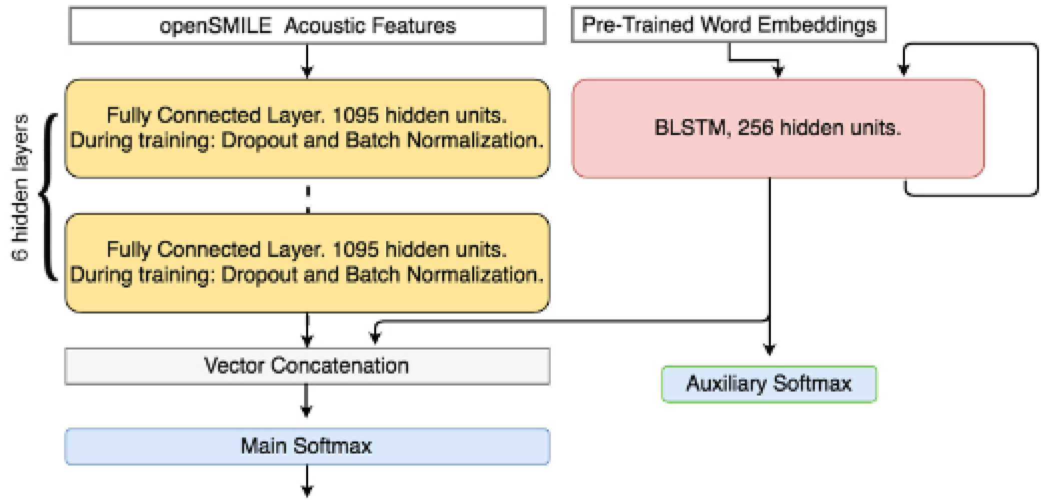

Figure 1: *Hybrid Acoustical Lexical Model Architecture, 4.2.5*

Figure 2. 8 BLSTMs for audio data

This is the architecture used to improve the f1 score of the model.

Table 4: *Classification experiments results (WE=Word Embeddings)*

Model	Features	Prec.	Recall	F1
LR	Trigrams	58.67	**63.95**	61.19
RF	OpenSMILE09	72.67	50.44	59.54
RF	OpenSMILE09, Trigrams	**76.11**	46.99	58.10
DNN	OpenSMILE13	63.65	58.03	60.71
DNN	OpenSMILE09	65.87	59.84	62.71
BLSTM	MFCC	54.19	55.10	54.64
BLSTM	WE	60.46	60.45	60.46
Hybrid	OpenSMILE09, WE	67.32	60.80	**63.90**

Figure 2. 9 Results for ML, BLSTM and hybrid models

The BLSTM model is a modification of the LSTM model. It analyzes input simultaneously in the forward and reverse time directions. Both models' effectiveness come from the LSTM node's capacity to retain memory of its prior values with an internal state. This bridges long temporal gaps.

2.4. Video analysis literature

2.4.1. Using Local Binary pattern and gaze features:

[8] Detection creates and analyzes a video dataset where each participant is shown 6-10 images from the selected image set, one at a time. (S)he is asked to describe the image on the screen. The participant is free to describe the image honestly or deceptively.

Modalities:

1. Video : Local Binary Pattern (LBP) features
2. Audio : zero crossing rate, spectral centroid, spectral bandwidth, spectral rolloff, chroma frequencies and mel frequency cepstral coefficients (MFCC)
3. EEG electroencephalogram readings (invasive)
4. Eye Gaze : uses infrared light to track data points like a user's pupil dilation, gaze movement, gaze direction, gaze fixations (pauses and lingers, useful to track pattern and focus), left pupil size, andright pupil size – PyGaze analysis library

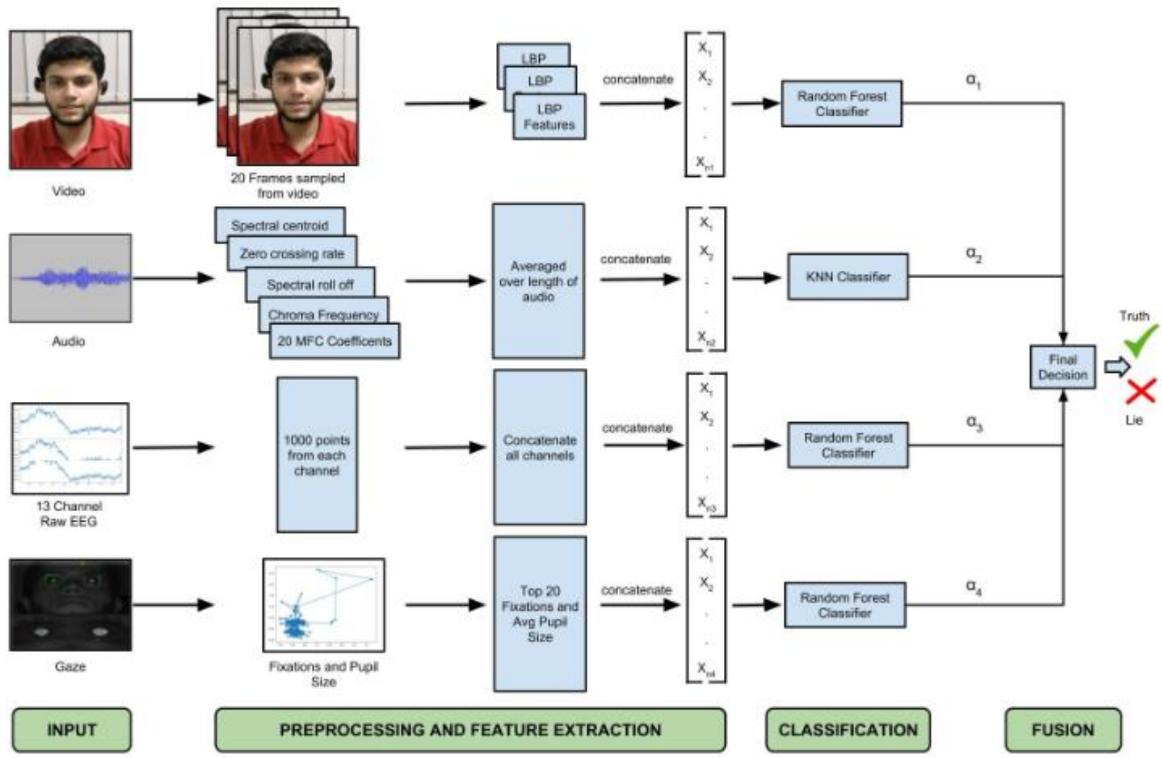

Figure 2. 10 Architecture for multi-modal analysis using text, gaze, LBP, EEG inputs

Modality	Method	Average Accuracy	
		Set A	Set B
Only EEG	Random Forest	58.71	-
	EEG Net	54.25	-
	MLP	53.79	-
Only Gaze	Random Forest	61.70	57.11
	MLP	57.71	53.51
Only Video	LBP + SVM	55.21	53.25
	LBP + Random Forest	56.20	55.26
	LBP + MLP	54.22	49.90
Only Audio	Random Forest	53.24	54.89
	KNN	53.22	56.22
EEG + Gaze	Score level fusion of best performing algorithms on various modalities	62.22	-
EEG + Audio		61.69	-
EEG + Video		60.20	-
Gaze + Audio		63.69	59.42
Gaze + Video		62.19	62.71
Audio + Video		60.68	58.24
Gaze + Video + EEG		62.70	-
Gaze + Audio + EEG		63.21	-
Audio + Video + EEG		63.18	-
Gaze + Video + Audio		64.69	60.09
All four		66.17	-

Figure 2. 11 Results for Bag-of-Lies dataset

2.4.2.Using Haar-cascade facial features and Silent Talker algorithm

The paper [14] creates and analyzes a dataset with a holiday-inspired role-playing exercise while they are being asked questions about items they packed in their suit-case. It extracts facial and eye micro movements using Silent Talker Algorithm and Haar Cascade features. The most successful classifier (Random Forest) achieved an overall classification accuracy of 78% in 10 recursive runs while trained over 80 subjects

Algorithm 1. Channel data extraction from video files using Silent Talker and Haar Cascade

Inputs: Video data stream *(ν)*
Output: Vector □: extracted Cannel data
Procedure:
Set *s-index* to 1st video frame in *ν*
Step 1: Take one Slot (*ș*: 1 sec) of *ν*
For-each video frame (*f*) in *ș*
Search for '*face*' using **Haar** Cascade:
If a 'face' is identified
F_c ← Rectangular coordinates around the 'face'
Search for 'eyes' within F_c using **Haar** Cascade:
If two 'eyes' are identified
L_c ← Left eye coordinates
R_c ← Right eye coordinates
Obj[] ← ST: **Object Locator(*F_c, L_c, R_c*)**
goodFrame ← ST: **is_Good_Frame(*Obj[],f*)**
if (***goodFrame***)
□ ← ST: **Channel Coder(*obj[], f*)**
End loop
Increase *ș-index* by 1 to get next overlapped slot *ș*
Go to Step 1 until last *ș* in *ν*

Figure 2. 12 Algorithm for video channel data extraction using Haar Cascade, Silent Talker

Table 2
Non-Verbal Channel list extracted from the video dataset using algorithm 1.

Channel NO	Channel Name	Channel Category	Channel No	Channel Name	Channel Category
1	face vertical movement (fvm)	face	19	left eye shift (lshift)	eyes
2	face horizontal movement (fhm)	face	20	left eye closed (lclosed)	eyes
3	face scale (fs) change(forward/backward movement)	face	21	left eye half left (lhleft)	eyes
4	face blush (fblu)	face	22	left eye half right (lhright)	eyes
5	face blanch (fbla)	face	23	left eye half closed (lhclosed)	eyes
6	face upward movement (fum)	face	24	right eye blink (rblink)	eyes
7	face downward movement (fdm)	face	25	right eye left (rleft)	eyes
8	face left movement (flm)	face	26	right eye right (rright)	eyes
9	face right movement (frm)	face	27	right eye shift (rshift)	eyes
10	face forward movement (ffm)	face	28	right eye closed (rclosed)	eyes
11	face backward movement (ffm)	face	28	right eye half left (rhleft)	eyes
12	face vertical shift (fvs)	face	30	right eye half right (rhright)	eyes
13	face horizontal shift (fhs)	face	31	right eye half closed (rhclosed)	eyes
14	face vertical shift with noise (fvsn)	face	32	face movement clockwise (fmc)	face angle
15	face horizontal shift with noise (fhsn)	face	33	face movement anti-clockwise (fmac)	face angle
16	left eye blink (lblink)	eyes	34	face movement angle-change (fma)	face angle
17	left eye left (lleft)	eyes	35	face movement right (fmuor)	face angle
18	left eye right (lright)	eyes	36	face movement left (fmuol)	face angle

Figure 2. 13 Facial and eye micro movement features – Silent Talker

2.4.3.Using local facial and muscle movement features

This paper [10] analyzes the real world dataset of court trials by extracting the following features

Local facial features/muscle movements (eg jaw contraction)

1. A dense representation, using HOG and LBP descriptors

2. A trajectory description, using IDT algorithm
3. a spatio-temporal characterization, using LBP-TOP
4. and a points of interest depiction with respect to agiven face, previously identified via OpenFace, using ORB feature extraction

2.5.Personality factors literature

The Five Factor Model assumes that our personality can be defined by the following five factors.

Personality Trait	Description	Comparison
Extroversion (E)	The personality trait of seeking fulfillment from sources outside the self orin community. High scorers tend to be very social while low scorers prefer to work on theirprojects alone.	reserved, thoughtful vs sociable, fun-loving
Agreeableness (A)	Reflects much individuals adjust their behavior to suit others. High scorersare typically polite and like people. Low scorers tend to 'tell it like it is'.	suspicious, uncooperative vs trusting, helpful
Conscientiousness (C)	The personality trait of being honest and hardworking. High scorerstend to follow rules and prefer clean homes. Low scorers may be messy and cheat others.	impulsive, disorganised vs disciplined, careful
Neuroticism (N)	The personality trait of being emotional.	calm, confident vs anxious pessimistic
Openness to Experience (O)	The personality trait of seeking new experience and intellectual pursuits. High scores may day dream a lot. Low scorers may be very down to earth.	prefers routine, practical vs imaginative, spontaneous

Table 2. 5 Five factor personality factors

[19] analyzed a corpus where participants filled the Ten-Item Personality Inventory, and the Rosen-berg Self-Esteem Scale and a survey questionnaire about lying. Lies were categorized into three types : altruistic, self-serving,and vindictive.

Table 4 Standard multiple regression analyses for lying and correlation analyses

Lie type and predictors	R	R^2	Adj. R^2	F	B	β	sr^2	r
Self-serving	.48	.23	.22	17.01***				
Self-esteem					−.08	−.35***	.07	−.44***
Openness					−.05	−.09	.01	−.18***
Conscientiousness					−.06	−.14	.00	−.31***
Extraversion					.00	.01	.00	−.17**
Agreeableness					.04	.09	.01	−.19***
Neuroticism					.00	.00	.00	.26***
Altruistic	.23	.06	.14	3.32**				
Self-esteem					−.05	−.24***	.03	−.21***
Openness					−.03	−.06	.00	−.08
Conscientiousness					−.02	−.04	.00	−.12*
Extraversion					.02	.05	.00	−.04
Agreeableness					.02	−.05	.00	−.01
Neuroticism					−.02	−.05	.00	.08
Vindictive	.33	.11	.09	6.86***				
Self-esteem					−.01	−.09	.00	−.05
Openness					−.01	−.02	.00	−.05
Conscientiousness					−.01	−.02	.00	−.08
Extraversion					.03	.10	.01	.04
Agreeableness					−.11	−.32***	.09	−.30***
Neuroticism					−.04	−.13*	.01	−.01

*$p < .05$
**$p < .01$
***$p < .001$

Figure 2. 14 Big Five Personality correlations with self reporting lie questionnaire

Neuroticism was positively correlated with self-serving lies. For self-serving lies and altruistic lies, self-esteem was a significant predictor variable.

3. : Problem Statement and Approach

3.1.Problem Description

We propose to create an automatic deception detection (ADD) model using the verbal and behavioral input of the person of interest.

Our objective is multifold:

1. To generate a binary output – Truth or Lie, for a given input of certain formats.
2. To understand the features that contribute to the truthfulness of a statement.
    a. Textual features
    b. Audio features
    c. Facial features
3. To understand what kind of questions trigger responses that are easily identifiable as a lie.
4. To check and understand if and how personality contributes to the ability to deceive.
5. To evaluate the efficacy of applying the state-of-the-art text, speech and video processing techniques to the problem of automatic deception detection.

3.2.Corpus

An essential contribution of this work is the creation of a deception corpus. The ideal data set attributes were taken into account before the collection of this Corpus. We needed a cleanly recorded speech and video data set for which ground truth is known with certainty and balanced output classes. We designed the data collection keeping these points in mind.

3.2.1. Method

3.2.1.1.Flowchart

The following is a flowchart for all the steps for the project.

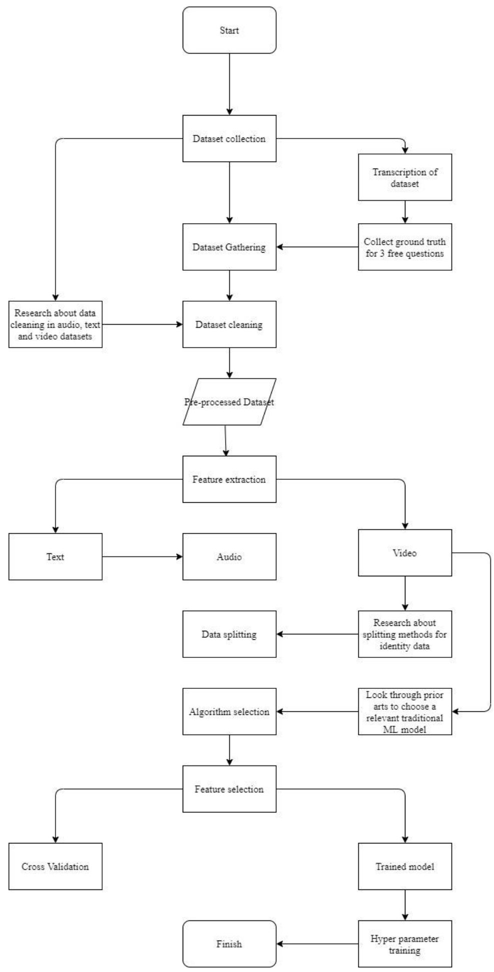

Figure 3. 1 Flowchart of methodology

3.2.1.2. Dataset description

The data set is based on a fake resume deception paradigm. The participants were applicants in a simulated job interview and were asked to answer biographical questions. For each question, a display on the screen (invisible to the interviewer) instructs the participant about the manner in which they must answer the question. One of these three prompts is shown: Lie, truth, or impress. For each of the three types of answers, the candidate tries to create a good impression with the interviewer. When the screen displays a lie or truth, they must answer accordingly. The option 'Impress' allows the participant to answer the question in freestyle; that is, they are given an option to tell a lie. Their responses are monitored and recorded using two cameras, one focused on facial expression and the other observing their body gestures. A microphone records their speech.

Speakers were offered a monetary incentive for their time and participation.

The following are the components of the data collected:

1. Speech
2. Front camera recording (facial expressions)
3. Side camera recording (hand gestures and pose)
4. Big five personality test

The following are a few features of the data collected:

1. Two practice questions to understand the experiment setting.
2. Sentence wise truth identification for the freestyle questions
3. Control using within-subjects experiment design by making the participant answer each question truthfully and deceptively.
4. Two questions followed between subjects' experiment design, where the interviewer phrased the question differently for different candidates.
5. Different types of questions: Leading questions and open-ended questions

3.2.1.3.Types of questions asked

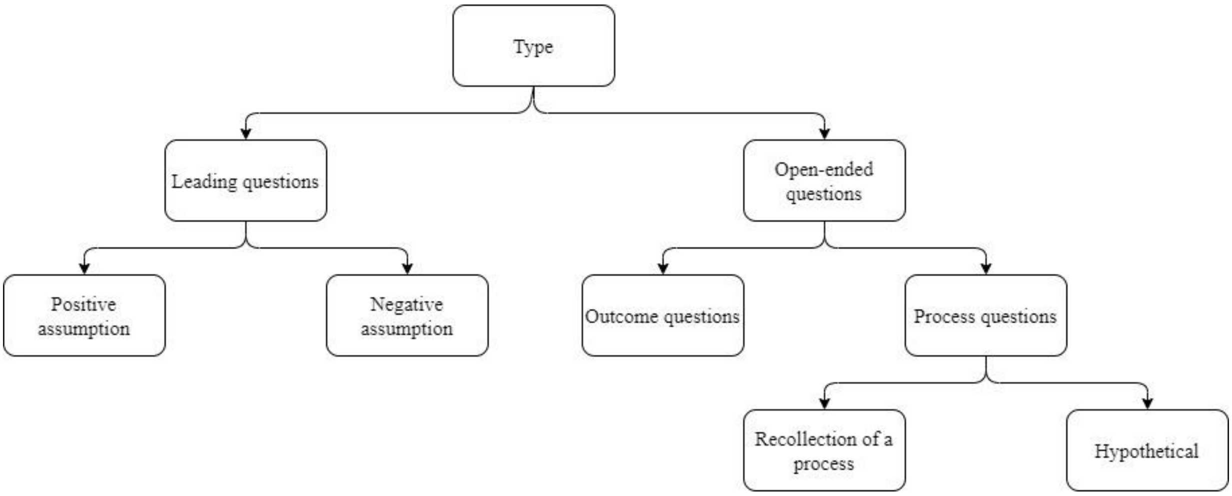

Figure 3. 2 Question categories

Different ways of formulating questions influence the veracity of responses. According to [20], negative assumption questions that presuppose a problem promote more frequent disclosure of undesirable work-related behaviors. Thus, we included two kinds of leading questions:

1. Positive assumption questions presuppose the absence of a problem. These questions indicate that the speaker is aware of a particular issue, but is averse to confrontation.
2. Negative assumption questions that presuppose a problem. Negative assumption questions are assertive even at the cost of potential discomfort.

General questions or open-ended questions create an impression that the interviewer lacks specific knowledge and thus, is unlikely to detect any lies. We have classified open-ended questions into 2 categories according to [21].

1. Outcome questions. These questions can be anticipated and prepared for. Examples include reasons for travel, number of years of work experience, etc.

2. Process questions. These depend on the progressional experience of an event. Truth tellers who have experienced a particular event draw upon the memory to answer questions related to the processes involved in the act. Examples include, what are the challenges faced while learning a concept.

The process questions were classified into two categories:

1. Hypothetical questions: These questions required the participant to understand a hypothetical scenario and answer accordingly. This does not need the person to recollect a memory or past experience. We expect that the lies and truth to these questions will be similar.
2. Recollection questions: A good storyteller can make an answer on the fly quickly when asked to lie. A truth teller will need to do a memory search to find the best yet truthful answer. Most recollection processes are triggered when asked about a recent time-period. Questions about significant life events have practiced answers ready, whereas recent events need a moment of thought.

For the three questions, where we asked the participant to answer freestyle, giving them a choice to lie, we later collected the veracity of each sentence from the participants. For each question, a different threshold was set to categorize that response into truth or lie, depending on the length of the answers. For questions that trigger short responses, a larger threshold is set to categorize the answer as truthful. Even if one out of three sentences is a lie, the answer is considered a lie. Even when we lie to questions in real life, it is difficult to conjure an entirely different situation. In general, 70% of a deceptive conversation is based on true occurrences. We hide small lies in a blanket of truth. The following are the threshold for the three questions: Question 4: 0.7, Question 7: 0.85, Question 15: 0.8

3.2.2.Example question and answer:

Example Question	How did you come to acquire the knowledge required for this role?
Example Lie Answer	I was interested in this since childhood so ever since my 7th standard I'd already started learning about calculus and relevant stuff, and I think in my 10th or 11th I'd already completed the statistics course of some undergraduate math degree. So I think ever since I can remember after my 6 or 7 standard I've been pursuing this as my primary.
Corresponding Truth Answer	I've been learning this since a couple of years since my college times, so once I realized that I do not want to continue practicing as a clinical doctor, I immediately started finding out alternative ways of learning, and I realized that statistical fields of research are very interesting to me, so I started learning them, and for the next couple of years at least, I was just learning these things. Almost full time. And after that I actually ended up doing projects which are also a great way to learn the skills required to become. Skills that I consider one needs to be.

Figure 3. 3 Example frames of the video dataset

3.2.3.Recording and labelling:

The data was segmented by splitting at the silences in the audio. The transcripts were generated using Microsoft Azure speech-to-text Batch transcription. The transcriptions were manually labeled as responses for different questions.

3.2.4.Challenges faced:

This experiment was conducted through an online medium on a video call to maintain social distancing norms while facing a sudden rise in Covid-19 cases in India. This method has multiple advantages such as:

1. Ease of recording and data-sharing.
2. Access to people from various areas with internet connectivity and laptop being the only requirements.
3. A new understanding of truth-lie behaviors in a modern world setup.
4. A novel understanding of truth-lie behaviors in a known and comfortable surrounding for the participant.
5. The video and audio quality were not optimal due to the internet issues and laptop specs on the participants' ends. However, this made our data set close to a real-world scenario with minimal equipment and can be used to train a model for the domain of online interviews.

A few disadvantages, such as the absence of an actual human being in front of the participant, were countered by maintaining a continuous video call of the interviewer so that the participants had someone to look at other than their blank laptop screens or their own video.

3.2.5.Data cleaning:

The generated transcript needed a significant amount of manual labeling and data cleaning. The chunks of dialogue generated between silences required to be merged to form answers to each question. Two responses by ID6 needed to be deleted due to high background noise.

3.2.6.Data splitting:

Data overlap when data points belonging to the same Id (participant) occur in more than one data split. For example, if a person's data points are present in the training and validation data set, then the model Learns to perform well on specific features. Our objective is to generalize the learning to all individuals. Hence, the data split was done so that there is no overlap between data points of a person in more than one data split.

The textual and audio features were split into two parts: Training 80% and test 20%.

The video features were split into three parts: Training 70%, validation 20%, and test 10%.

3.2.7.Feature selection:

The following methods were explored while selecting the features for text data

3.2.7.1.correlation features

The features that have high correlation with the output variable were considered important features. Correlation function was computed between each feature and the output variable to determine 150 important features for each category of text features (tF idf, lwc, POS tagging).

3.2.7.2.SelectKBest feature selection:

Using select K best feature selection technique, the Chi square, Logistic Regression, Random Forest and LGBM classifiers are used to select important features from the data set. Each set of selected features is modelled separately and the best models are presented in the Results Section.

4. Modeling and Implementation

4.1.Text data

There are four categories of feature classes that are extracted from this data set.

4.1.1.Data set:

The text data set is the set of transcripts generated for each question. The average word count for each response is 85. There are 897 entries. Each of the 31 participants was asked 29 questions.

4.1.2.Feature extraction

4.1.2.1.Tf-idf features:

Tf-idf feature unigrams bigrams and trigrams

It stands for term frequency-inverse document frequency. Term frequency refers to the number of occurrences of a given word in the document. If a word frequently occurs in a document, then that document is considered to be relevant for that word. We cannot use term frequency value alone; commonly occurring words such as articles might get a higher weightage since they occur in most documents.

Inverse document frequency. For the presence of a word among all documents, if a word occurs only in a few documents, then it's given a higher ideal value. Thus, rare words in a corpus have a high idf value. Through this, infrequent essential words are highlighted and frequent non-useful words such as articles and prepositions are penalized by the idf value.

Tf-idf is the product of term frequency and inverse document frequency. It gives a score to documents based on both statistics. Thus, it tells you the importance of a given word in a delivered document with respect to the whole Corpus.

4.1.2.1.1.Model evaluation

There were 42662 features since the data set was extracted for unigrams, bigrams and trigrams. Vectorisation is done on the extracted features.Correlation feature selection was applied to determine 100 features which correlated well with the output variable.

Dataset: 897 x 42,662

Feature selection: correlation feature selection

After feature selection : 897 x 100

Best algorithm: K nearest neighbours model

Hyperparameter tuning result: Best leaf size: 1, best p:1, best n neighbours: 3

Train accuracy : 64%

Test accuracy: 62 %

4.1.2.2.LIWC features:

LIWC is a popular text analysis tool that calculates word counts for 72 linguistic dimensions. It can assist judgement by automating the assessment of large volumes of responses for deception and guiding human decision-making regarding the veracity of a statement. We extracted 93 features for our data set. However, since ours is a transcribed dataset, we had to remove four features: Semicolon, exclamation, quote and parentheses, which had zero value for all entries.

Following are important categories in LIWC:

NP model: Cognitive processes : affect, time, see, hear, space

RM approach: First person singular, pronoun, negative emotions, conjunction, motion verbs

4.1.2.2.1.Model evaluation:

Dataset: 897 x 94

Best algorithm: Linear Support Vector Machine

Hyperparameter tuning result: C= 1, gamma = 1, kernel = Linear

Train accuracy : 59%

Test accuracy: 57 %

Consistent with the reality monitoring Framework model, cognitive processes like time and space are significant features. Consistent with the NP model, pronouns and negative emotions are key features. Hedge features like tentative words are also in the top 15 important features.

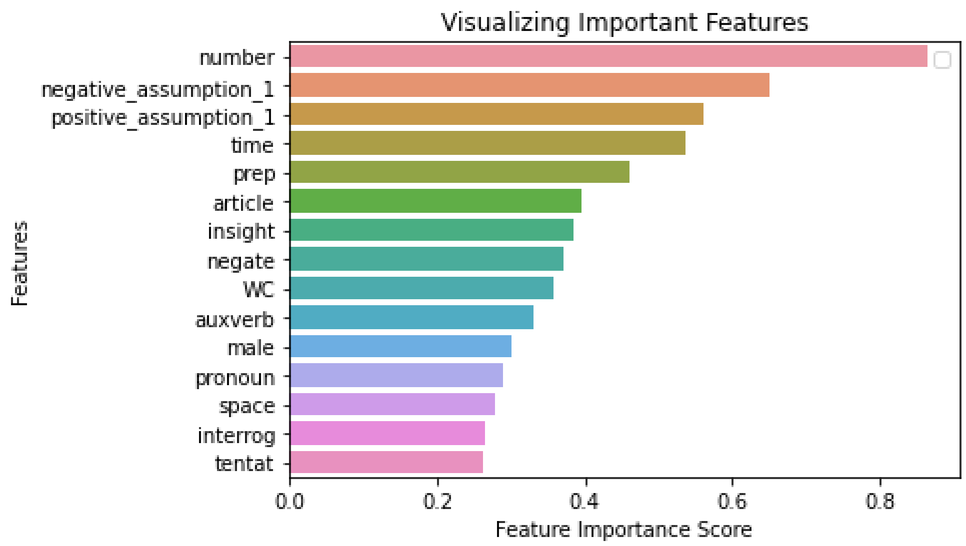

Figure 4. 1 LIWC - Top 15 features for LinearSVC Model

4.1.2.3.POS taggers

POS tagging is the process of marking a word to a corresponding part-of-speech tag based on the definition and its context. The context is essential since the same word can be tagged as different parts of speech based on its usage in the sentence. For example, in the sentence 'Walk with me.', 'walk' is a verb but, in the sentence, 'Let's go for a walk.', 'work' is a noun.

We used pre-trained a classifier-based POS tagger on the penn treebank Corpus in nltk library for part of speech tagging.

Features are extracted from the words using pattern identification from regular expressions and unigrams, bigrams and trigrams by the feature detector and then are passed to an internal classifier. It classifies these features and returns the label.

4.1.2.3.1.Model evaluation:

Dataset: 897 x 94

Best algorithm: KNN classifier

Hyperparameter tuning result: k= 14 leaf_size = 30, p=2

Train accuracy : 56%

Test accuracy: 57 %

4.2.Audio data

4.2.1.Feature extraction

Literature suggests that pitch, energy, speak-ing rate, and other stylistic factors (e.g. "muffled" voice) vary when speakers attempt to deceive. We consider a good range of acoustic and prosodic features as a part of the ComPARE feature set. We extract these features using the OpenSMILE library. We divide the audio into question answer units and extract the features of every response. The features include durational, pausing, intonational, and loudness. Features are automatically normalized by the library, taking into consideration segmental context also as long-term speaker-specific habits.

4.2.1.1.Model evaluation:

Dataset: 897 x 6,382

Feature selection method: ExtraTreeClassifier

After feature selection : 897 x 2677

Best algorithm: Decision Tree

Hyperparameter tuning result: criterion: ‘gini’, splitter: ‘best’

Test accuracy: 68%

Dataset: 897 x 6,382

Feature selection method: Correlation

After feature selection : 897 x 100

Best algorithm: Logistic classifier and linear support vector classifier

Hyperparameter tuning result: C= 0.1, penalty = ‘l2’, solver = ‘libLinear’

Test accuracy: 62%, 62%

4.2.1.2.Comments:

The features spectral entropy segment length, fundamental frequency f0 final sma kurtosis ,log hnr-sma centroid ,mfcc Max segment length, spectral flux quartile are common in both classifiers and are important features in both classifiers.

Support Vector Classifier feature importances:

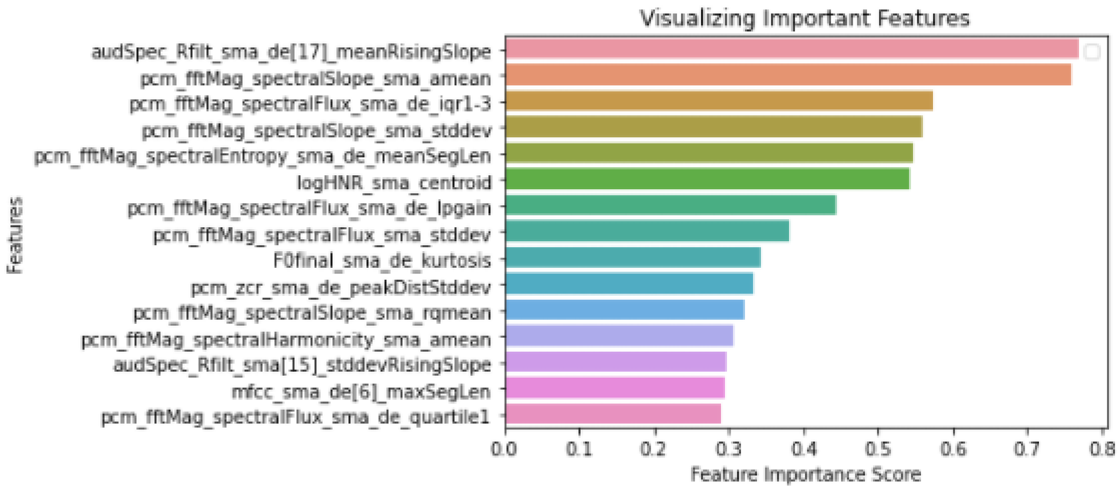

Figure 4. 2 Audio data LinearSVC - Top 15 features

Logistic Regression classifier feature importances:

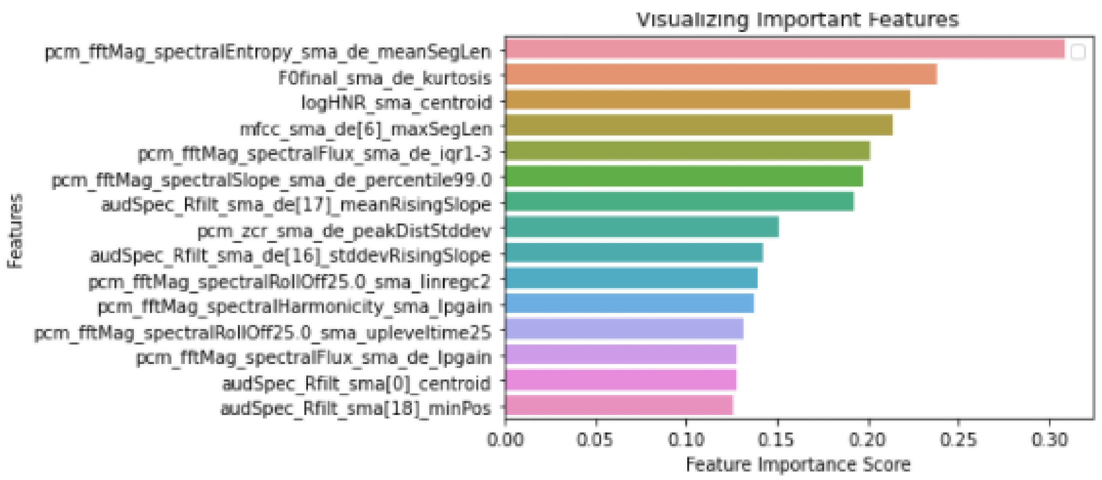

Figure 4. 3 Audio data : Logistic Regression - Top 15 features

4.3.Video data:

4.3.1. Feature extraction:

We use the OpenFace pretrained facial feature generating model for each frame of the video. During the training portion of the OpenFace pipeline, 500,000 images are passed through the neural net. OpenFace model trains these images to produce 128 facial embeddings which represent a generic face. It uses Google's FaceNet architecture for feature extraction and uses the triplet loss function for training.

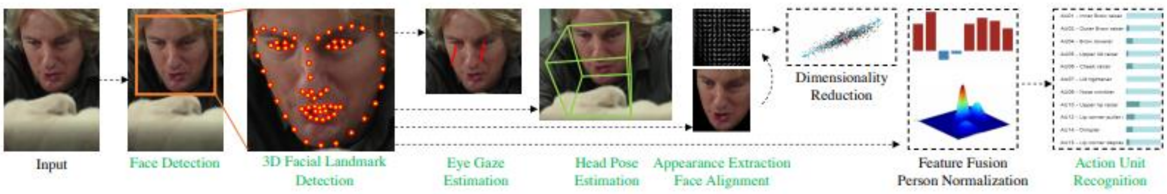

Figure 4. 4 OpenFACE pipeline for landmark detection, head pose and eye gaze estimation, and facial action unit recognition

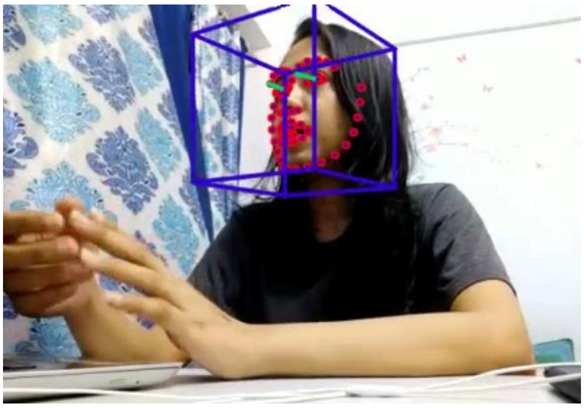

Figure 4. 5 OpenFACE eye gaze, head pose, landmark detection and facial unit recognition for an example from our dataset

4.3.2.Model fitting and evaluation:

After the open face features are extracted for each frame of the videos, an Lstm is fit to capture the sequential data in the frames. Categorical cross entropy is used as the loss function. The model is trained for 150 epochs with a total of 36,152 trainable parameters, yielding a test accuracy of 64.7 percent. The accuracy improved significantly for the training data while the validation set accuracy improves marginally. One hidden layer of 50 nodes along with a dropout of 0.5 followed by an output layer of 2 nodes is used. All the 720 features are inputted to get the best test accuracy.

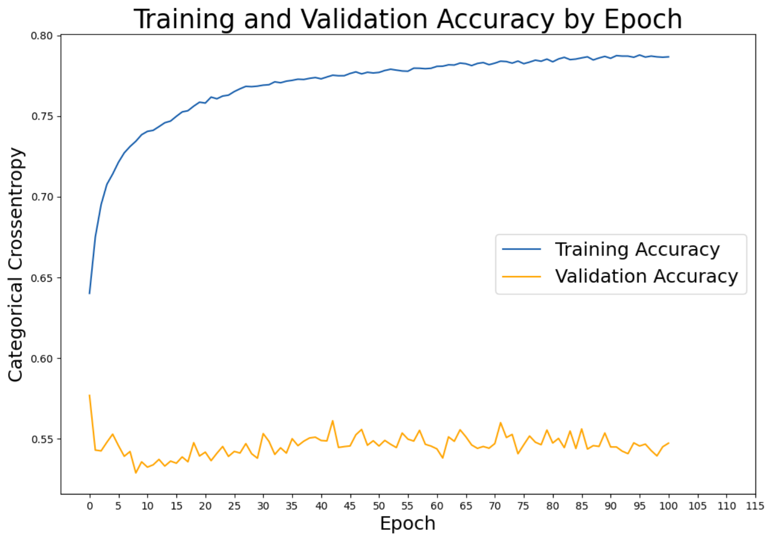

Figure 4. 6 Training plot for visual data

Layer (Type)	**Output Shape**	**Param #**
Dense	(720, 50)	36050
Dropout	(720, 50)	0
Dense 1	(50, 2)	102
Total Params	36152	
Trainable Params	36152	

Table 4. 1 LSTM model configuration

4.3.3. Results

	Features	**Feature Selection**	**Classifier**	**Accuracy**
Text	TF-IDF	Correlation	KNN	62%
	LIWC	Correlation	Linear SVC	57%
	POS	Correlation	KNN	57%
Audio	Prosodic, Acoustic and Deviation	Correlation	Linear SVC	62%
		Correlation	Logistic regression	62%
		Extratree classifier	Decision tree	68%
Video	OpenFace	All features	LSTM	64.70%
		Adaboost	LSTM	60.30%
		Correlation	LSTM	52.20%

Table 4. 2 Classification model results

5. : Summary and Conclusions

5.1.Summary

We have shown in this work that automatic features extracted through state of the art video and audio techniques are useful in the deception domain. We also show that video and audio can be automatically classified into deceptive and truthful parts with some degree of success. We have contributed a non native English speakers data set which is balanced in its lies and truth, and gender. It is unique as it contains various types of questioning, which can provide insight on which questions trigger responses that are easy to judge as deceptive . We have collected front and side camera videos for analysis of facial and hand gesture features. We have achieved an accuracy better than human baseline.

5.2. Conclusions

Traditional classification models work better than regression models on the textual data. However, the domain-specific words, such as ‘job’ and ‘projects’, are common in both truth and lies. This becomes noisy for the classifier. In the future work, domain specific words can be removed from the corpus. Logistic classifier and linear support vector classifier work best for audio data. Feature selection played a key role in audio data modelling, due to the large number of features. Features that are common between the best performing classifiers can be used as a control group.

5.3.Future Prospects

In the future work, we can

1. perform emotion detection in facial features.

2. perform group and question type dependent modelling by creating clusters and analysing each cluster with a different model.
3. split the data into sentence units rather than inter-pausal units and perform the entire analysis accordingly.
4. Use another text feature extraction software, named CohMetrix.
5. Extract other duration features for audio analysis.
6. Include the pose feature data into the dataset.

5.4.REFERENCES

1. Bond, C. F. (2008). Individual differences in judging deception: Accuracy and bias. Psychological Bulletin, 134(4), 477–492. doi:https://doi.org/10.1037/0033-2909.134.4.477
2. DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. Psychological Bulletin, (pp. 111(2), 203–243). doi:https://doi.org/10.1037/0033-2909.111.2.203
3. DePaulo, B. M. (1996). Lying in everyday life. Journal of Personality and Social Psychology, (pp. 70(5), 979–995). doi:https://doi.org/10.1037/0022-3514.70.5.979
4. Knapp, M. L. (1979). Telling it like it isn't: A review of theory and research.
5. Phillips, Michelle & Meek, Scott & Vendemia, Jennifer. (2011). Understanding the underlying structure of deceptive behaviors. . Personality and Individual Differences. doi:50. 783-789. 10.1016/j.paid.2010.12.031.]
6. Weiss, B. &. (2006). Looking Good and Lying to do it: Deception as an Impression Management Strategy in Job Interviews. Journal of Applied Social Psychology, 36(4), 1070–1086. doi: https://doi.org/10.1111/j.0021-9029.2006.00055.x
7. An, Guozhen & Levitan, Sarah & Hirschberg, Julia & Levitan, Rivka. (2018). Deep Personality Recognition for Deception Detection. 421-425. 10.21437/Interspeech.2018-2269.
8. V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh and M. Vatsa, "Bag-of-Lies: A Multimodal Dataset for Deception Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 83-90, doi: 10.1109/CVPRW.2019.00016.
9. Graciarena, M. & Shriberg, Elizabeth & Stolcke, A. & Enos, Frank & Hirschberg, Julia & Kajarekar, S.. (2006). Combining Prosodic Lexical and Cepstral Systems for Deceptive Speech Detection. 1. I - I. 10.1109/ICASSP.2006.1660200.

10. Avola D, Cascio M, Cinque L, Fagioli A, Foresti GL. LieToMe: An Ensemble Approach for Deception Detection from Facial Cues. Int J Neural Syst. 2021 Feb;31(2):2050068. doi: 10.1142/S0129065720500689. Epub 2020 Nov 16. PMID: 33200620.

11. Sarah I. Levitan, Guzhen An, Mandi Wang, Gideon Mendels, Julia Hirschberg, Michelle Levine, and Andrew Rosenberg. 2015. Cross-Cultural Production and Detection of Deception from Speech. In <i>Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection</i> (<i>WMDD '15</i>). Association for Computing Machinery, New York, NY, USA, 1–8. DOI:https://doi.org/10.1145/2823465.2823468

12. Forsyth, L, Anglim, J. Using text analysis software to detect deception in written short-answer questions in employee selection. Int J Select Assess. 2020; 28: 236–246. https://doi.org/10.1111/ijsa.12284

13. Kao, Yi-Ying & Chen, Po-Han & Tzeng, Chun-Chiao & Chen, Zi-Yuan & Shmueli, Boaz & Ku, Lun-Wei. (2020). Detecting Deceptive Language in Crime Interrogation. 10.1007/978-3-030-50341-3_7.

14. Khan, W, Crocket, K, Oshea, J, Hussain, A and Khan, B (2020) Deception in the Eyes of Deceiver: A Computer Vision and Machine Learning Based Automated Deception Detection. Expert Systems with Applications. ISSN 0957-4174

15. Miller, G. R., & Stiff, J. B. (1993). Deceptive communication. Newbury Park, CA: Sage.

16. Bachenko, Joan & Fitzpatrick, Eileen & Schonwetter, Michael. (2008). Verification and Implementation of Language-Based Deception Indicators in Civil and Criminal Narratives. Coling 2008 - 22nd International Conference on Computational Linguistics, Proceedings of the Conference. 1. 41-48. 10.3115/1599081.1599087.

17. Levitan, Sarah & An, Guozhen & Ma, Min & Levitan, Rivka & Rosenberg, Andrew & Hirschberg, Julia. (2016). Combining Acoustic-Prosodic, Lexical, and Phonotactic Features for Automatic Deception Detection. 10.21437/Interspeech.2016-1519.

18. Mendels, Gideon & Levitan, Sarah & Lee, Kai-Zhan & Hirschberg, Julia. (2017). Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection. 1472-1476. 10.21437/Interspeech.2017-1723.

19. Hart, Christian & Lemon, Rasheonia & Curtis, Drew & Griffith, James. (2020). Personality Traits Associated with Various Forms of Lying. Psychological Studies. 65. 10.1007/s12646-020-00563-x.

20. Minson, J., Eric M VanEpps, Jeremy A. Yip and M. Schweitzer. “Eliciting the truth, the whole truth, and nothing but the truth: The effect of question phrasing on

deception.” *Organizational Behavior and Human Decision Processes* 147 (2018): 76-93.

21. Jupe, LM, Vrij, A, Leal, S, Nahari, G. Are you for real? Exploring language use and unexpected process questions within the detection of identity deception. Appl Cognit Psychol. 2018; 32: 622– 634. https://doi.org/10.1002/acp.3446