

Deception Detection in Legal Settings: A Fusion-Based Model

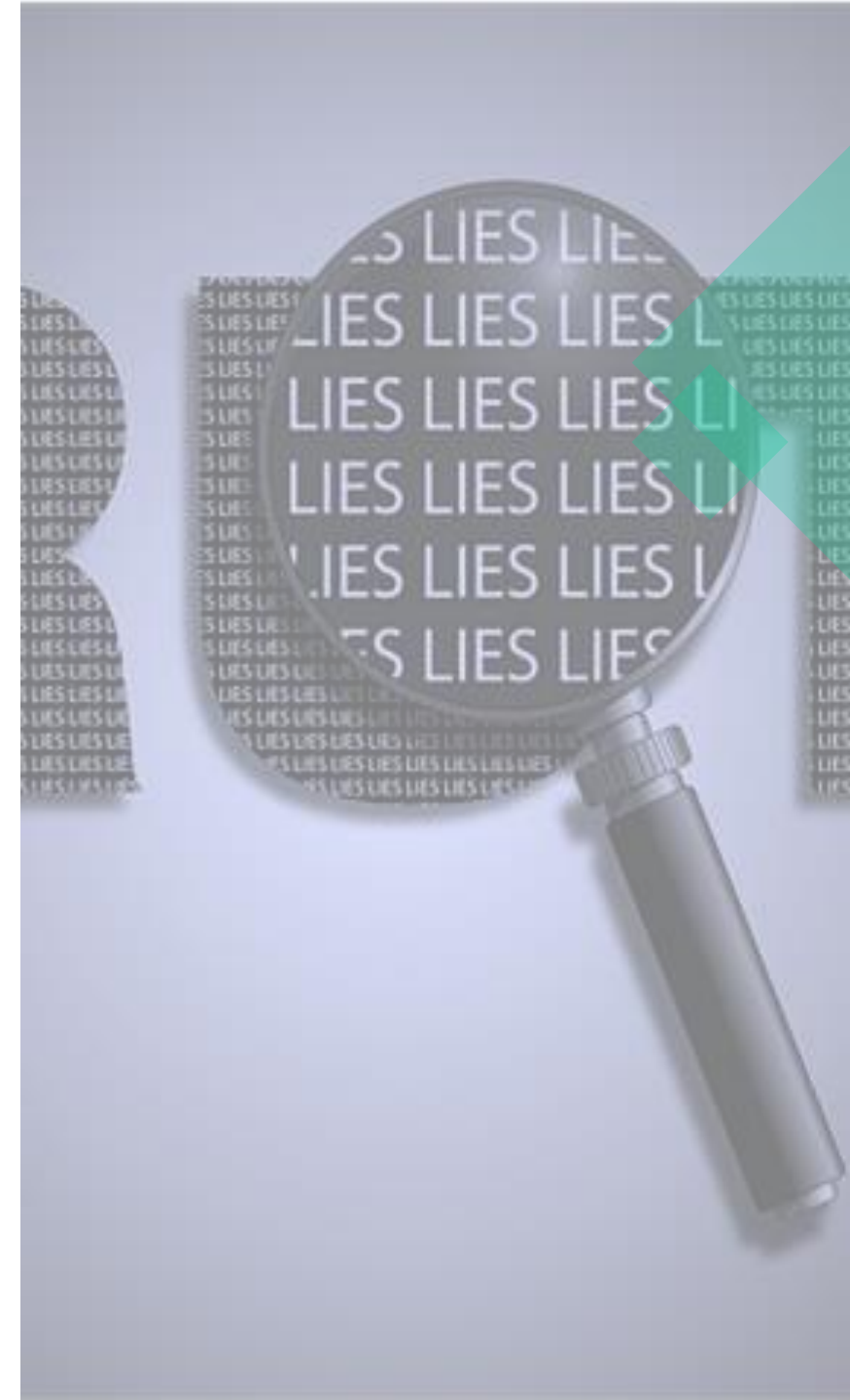
Technical Report





The problem

- **Deception** is the act of misleading or wrongly informing someone about the true nature of a situation
- Need to distinguish **deceptive** vs. **truthful** statements
 - Legal and law enforcement settings
 - Other domains (e.g. recruitment)
- **Poor detectors of deception**
 - Detection of lies no better than chance (Bond & DePaulo, 2006)
 - Holds across range of populations (e.g. job interviewers; law enforcement) personnel
 - Professionals never received formal training (Johnson, 2014)





Problem Statement/Aim

Can we accurately classify motivated **deceptive** vs. **truthful** responses through fusion of vocal, non-verbal, and lexical features?

Psychological theory was used to generate hypotheses and inform features selection

- **Newman-Pennebaker theory**

- Psychological distancing → Fewer first-person singular pronouns (e.g. 'I')
- Use of negative emotion words due to anxiety when lying results

- **Reality Monitoring**

- Lacking experiential information → less use of sensory information and object descriptions

- **Cognitive strain**

- Lying as cognitively taxing

Some hypotheses...

- Slower speech rates (vs. being truthful)
- Increased pauses and use of silence
- Increased use of negative emotion
- Decreased use of sensory-related words and nouns
- Decreased use of first personal singular pronouns



Data Set and Method





Original Data-set

- **Deceptive vs. Truthful Trial Testimonies/Statements**

- Witness/Defendant; interviews; media statements; exonerations

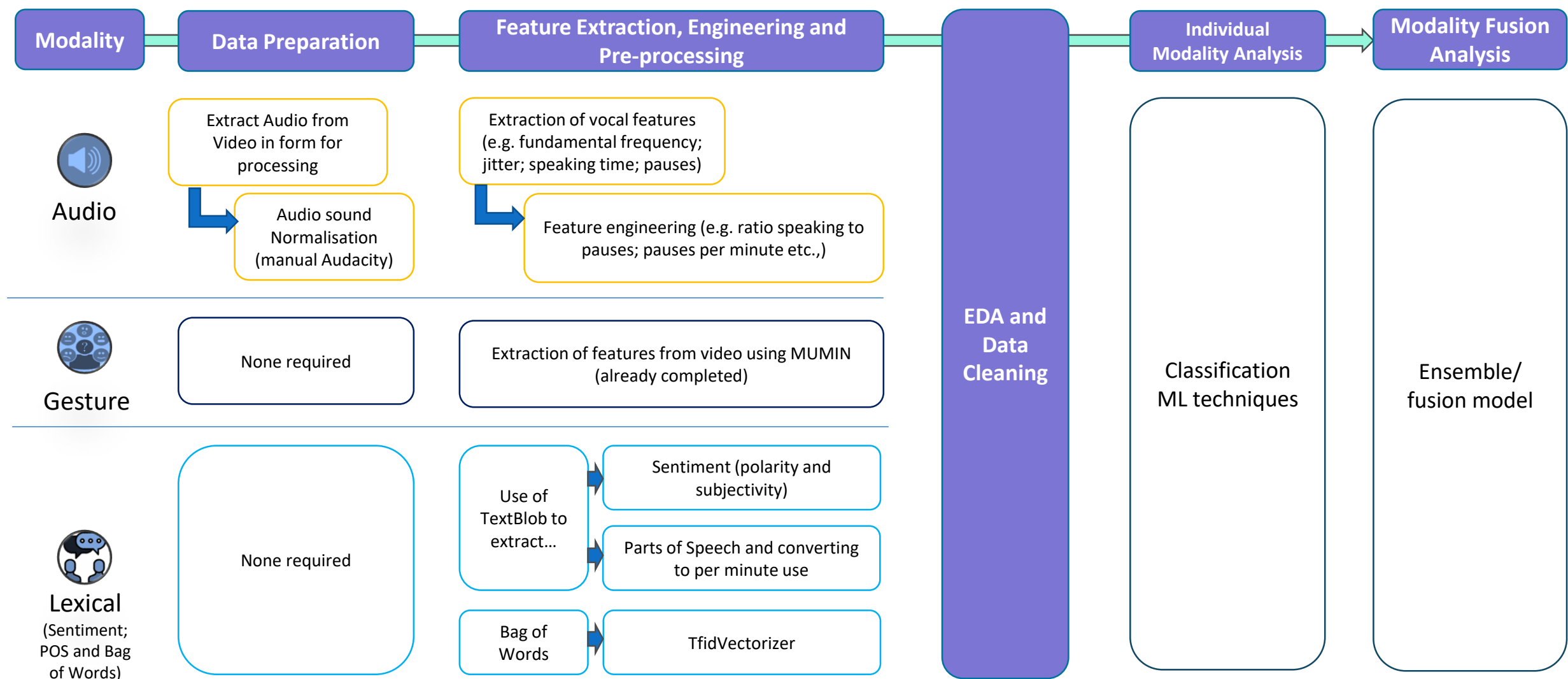
- **Multiple data forms**

- Videos¹
- Transcripts²
- Annotated gestures²

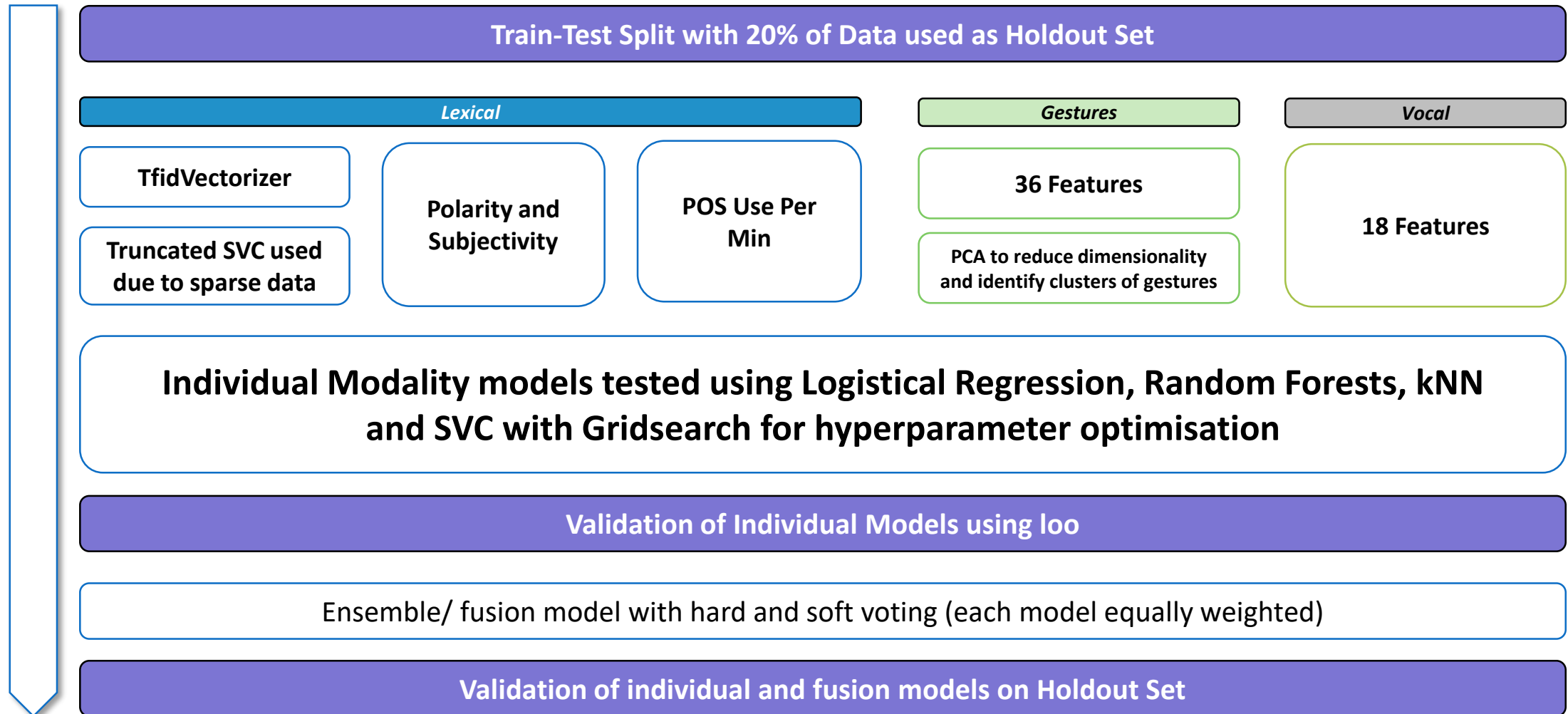
- **121 cases**


- 61 deceptive and 60 truthful videos
 - Trial outcomes as labels (guilty = deceptive; not guilty or exonerated = truthful)

Method: Features relevant to the five modalities were extracted using suitable Python libraries, with data cleaning and EDA then being undertaken prior to modelling



Analysis: Each individual model was assessed using Leave One Out (loo) on 80% of original data with 20% left out for validation. Model showing highest CV accuracy selected for fusion



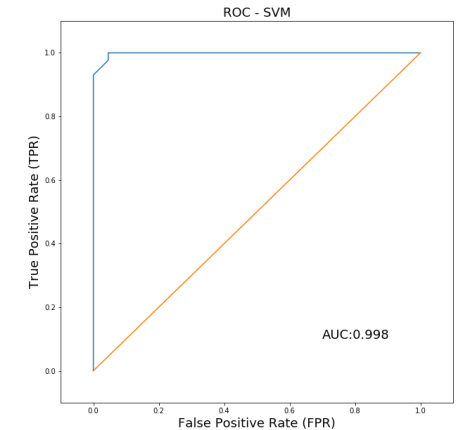
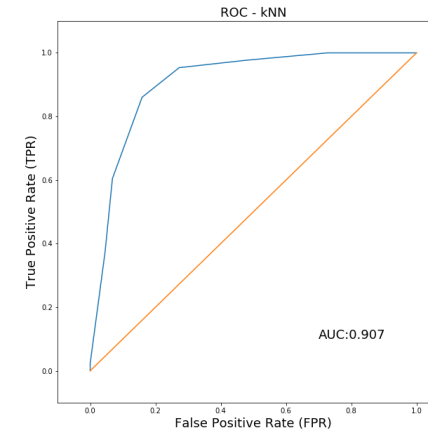
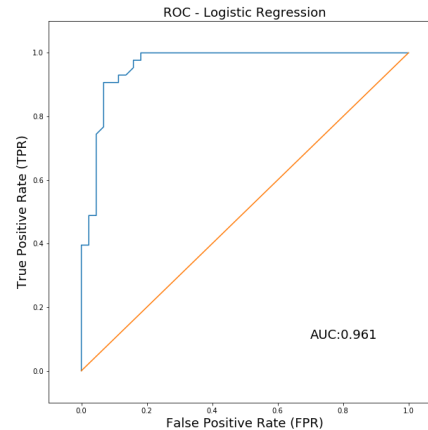
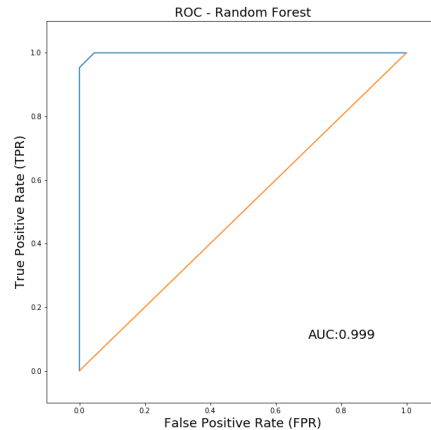


Results – Individual Models



Results

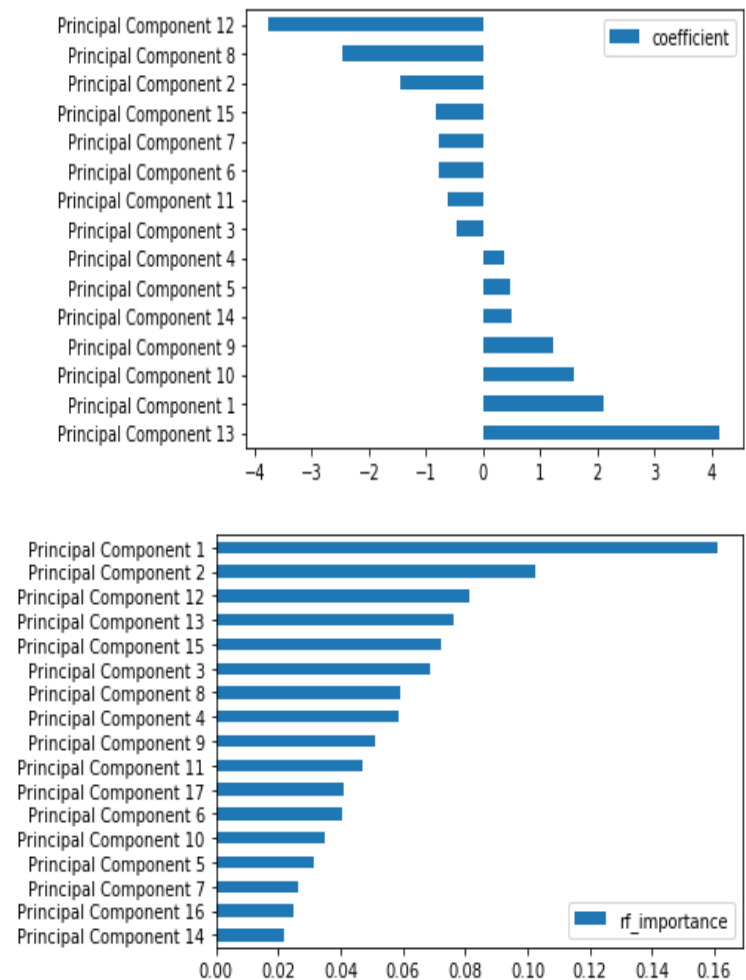
Gesture Results: The best overall model for gestures was SVM



	Random Forests	Logistical Regression	KNN	SVC
CV (Std)	0.74 (0.43)	0.77 (0.42)	0.77 (0.42)	0.78 (0.41)
F1	0.75	0.87	0.86	0.87
Best Parameters	<code>{'pca__n_components': 17, 'randomforestclassifier__criterion': 'gini', 'randomforestclassifier__max_depth': 6, 'randomforestclassifier__max_features': 5}</code>	<code>{'logisticregression__C': 10, 'logisticregression__penalty': 'l2', 'pca__n_components': 15}</code>	<code>{'kneighborsclassifier__algorithm': 'auto', 'kneighborsclassifier__n_neighbors': 7, 'kneighborsclassifier__p': 2, 'pca__n_components': 15}</code>	<code>{'pca__n_components': 17, 'svc__C': 10, 'svc__gamma': 0.0774263682681127, 'svc__kernel': 'rbf'}</code>



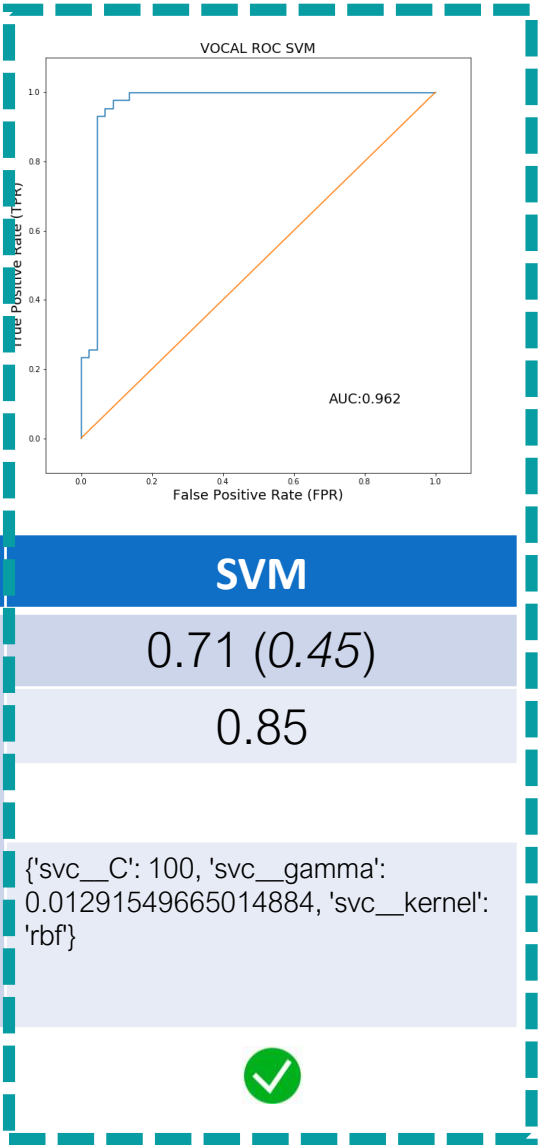
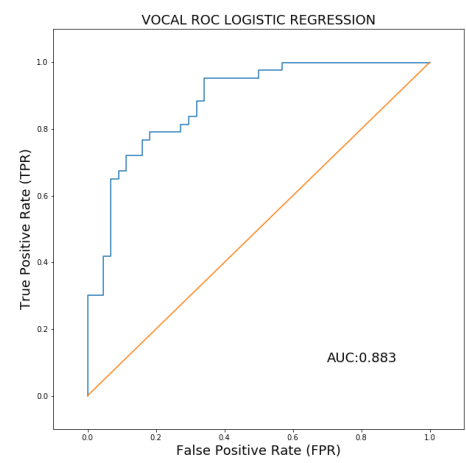
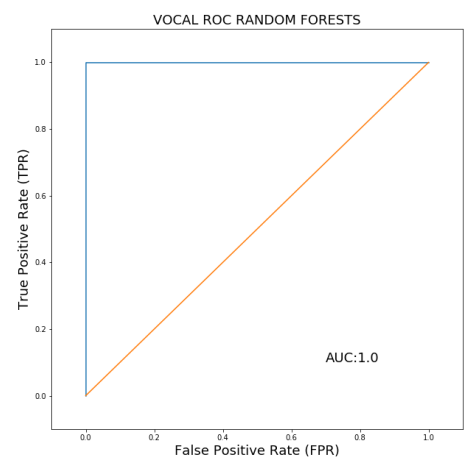
Difficult to assess feature importance in SVC due to rbf kernel but logistical regression and random forests indicating clusters of behaviours falling under components 1, 2, 12 and 13 best deception indicators



Gesture Clusters – Most Important

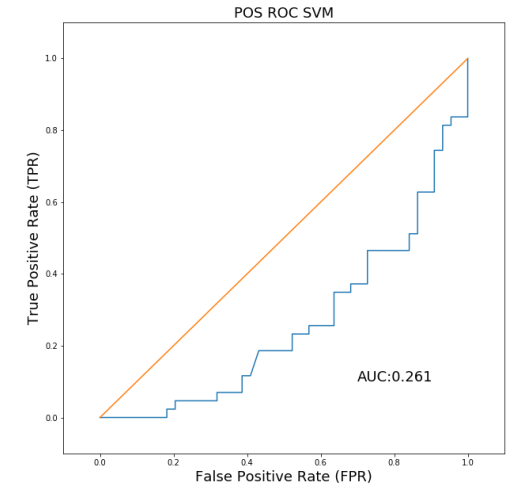
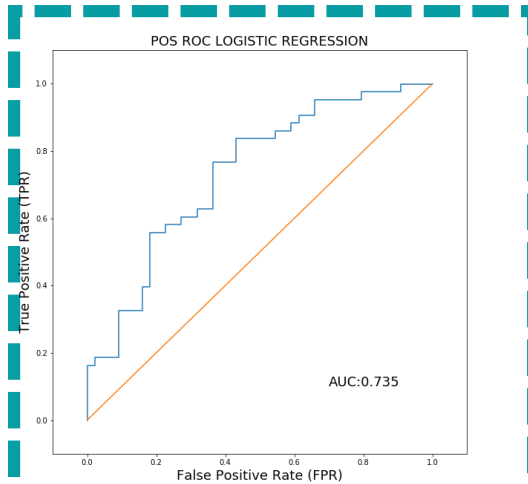
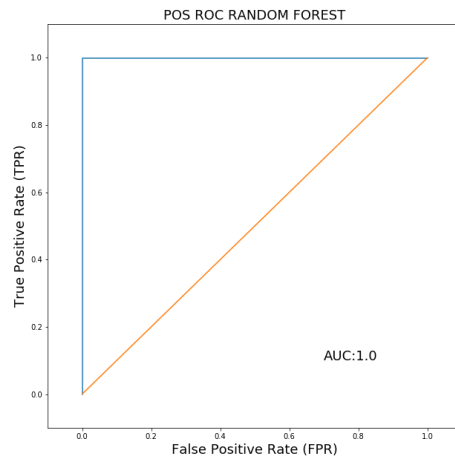
PC 1	PC2	PC 12	PC 13
Lips up at side	Rapid eye closing	Forward head tilt	Side tilt
Complex hand movements	Mouth opening	Less right forward head tilt	Less exaggerated eye opening
Use of both hands	Head tilted down to right	Less side turn to right	Less frowning

Vocal Results: The best overall model for vocal features was SVC



	Random Forests	Logistical Regression	KNN	SVM
CV (SD)	0.61 (0.49)	0.68 (0.47)	Not used	0.71 (0.45)
F1	0.80	0.84		0.85
Best Parameters	{'randomforestclassifier__criterion': 'gini', 'randomforestclassifier__max_depth': 5, 'randomforestclassifier__max_features': 4}	{'logisticregression__C': 10, 'logisticregression__penalty': 'l2', 'logisticregression__solver': 'newton-cg'}		{'svc__C': 100, 'svc__gamma': 0.01291549665014884, 'svc__kernel': 'rbf'}

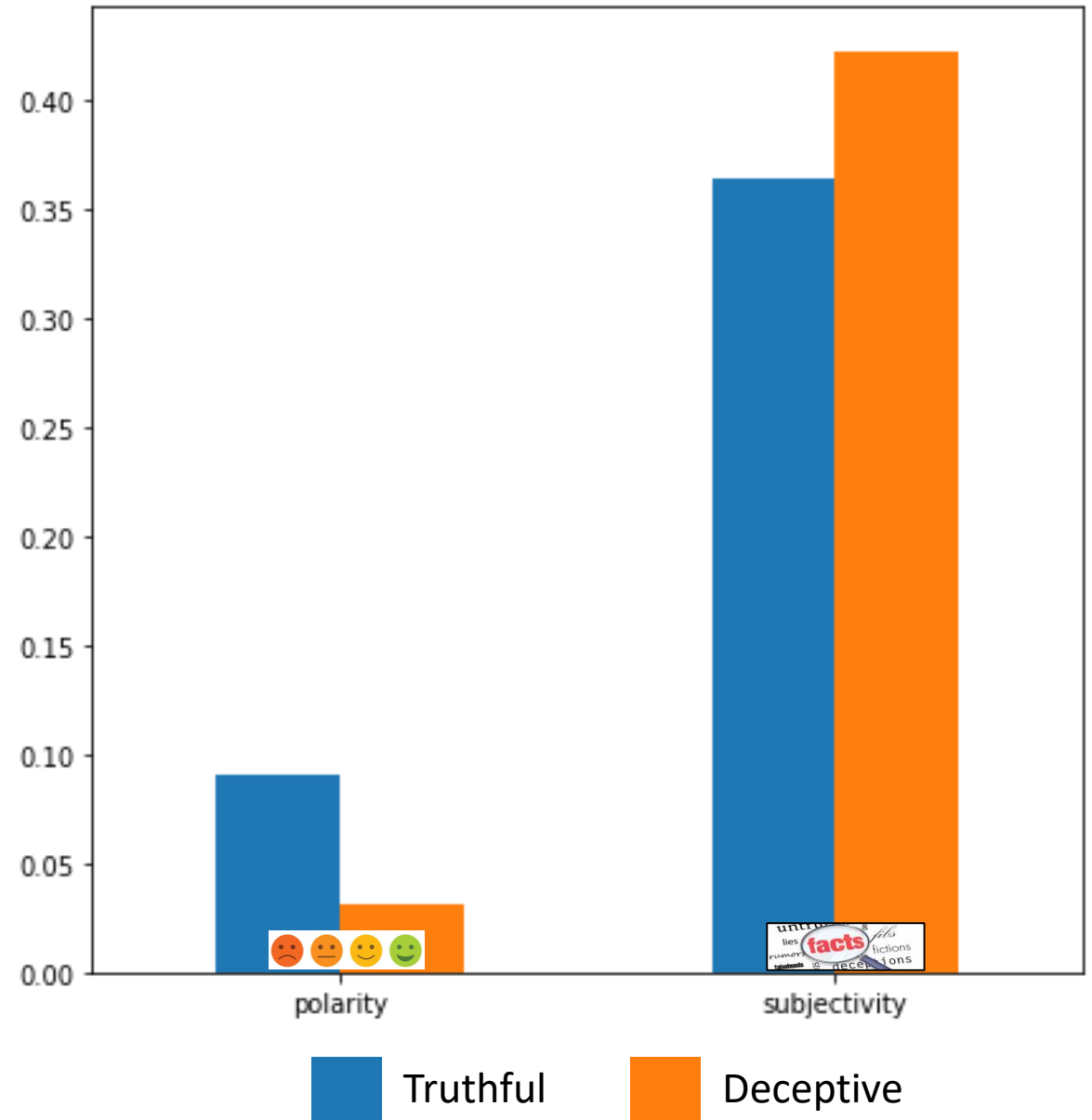
Parts of Speech Results: The best overall model for POS was Logistical Regression but all models were poor performing possibly due to issues around regularisation



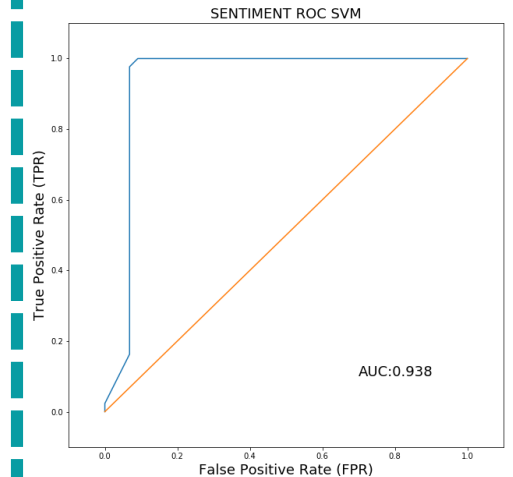
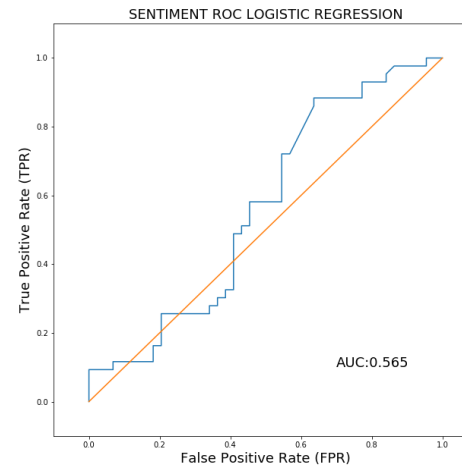
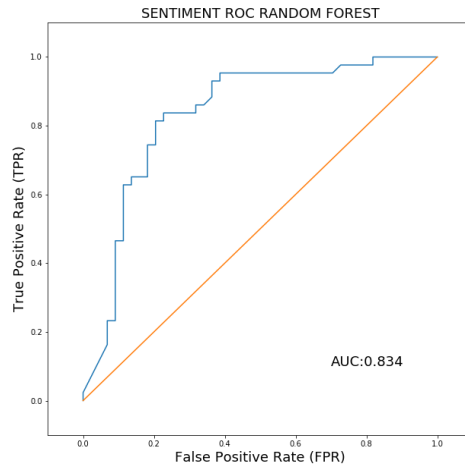
	Random Forests	Logistical Regression	KNN	SVC
CV (loo)	0.42 (0.49)	0.51 (0.50)		0.48 (0.50)
F1	0.69	0.78		0.85
Best Parameters	{'randomforestclassifier__criterion': 'entropy', 'randomforestclassifier__max_depth': 4, 'randomforestclassifier__max_features': 5}	{'logisticregression__C': 0.135, 'logisticregression__penalty': 'l2', 'logisticregression__solver': 'liblinear'}		{'svc__C': 0.17, 'svc__gamma': 1e-05, 'svc__kernel': 'linear'}



On average, **truthful statements** had **higher positive sentiment**. **Deceptive** statements though have **higher subjectivity** indicating higher levels of opinion, emotion, or judgement (vs. factual), aligning with theory



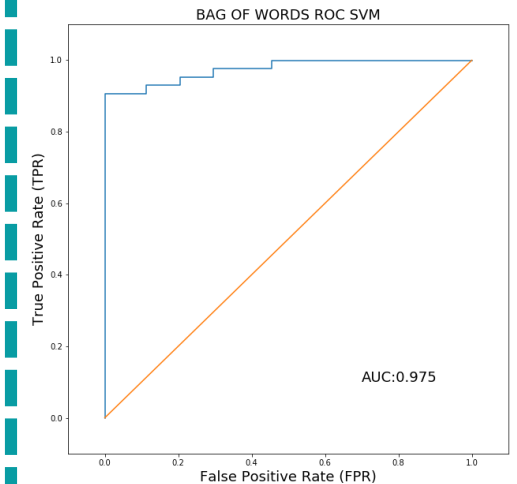
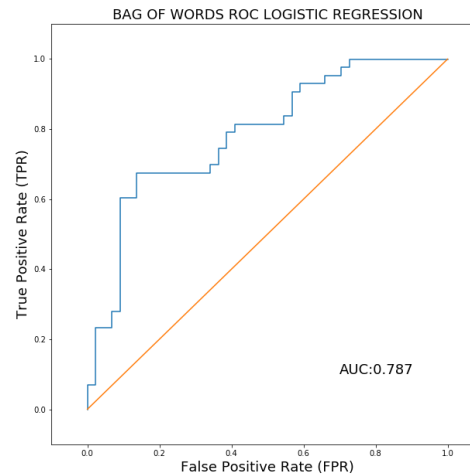
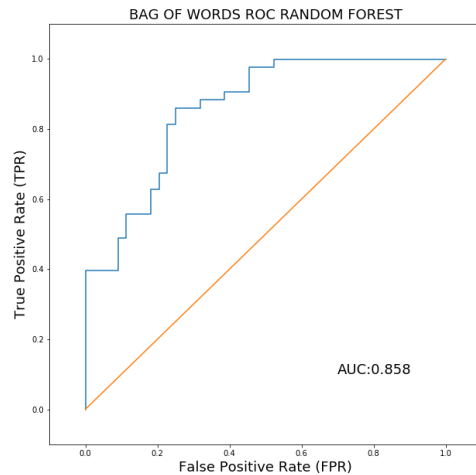
Sentiment Results: The best overall model for sentiment was SVC but important to recognise all models were poorly performing



	Random Forests	Logistical Regression	KNN	SVM
CV (Std)	0.59 (0.49)	0.49 (0.50)		0.66 (0.48)
F1	0.79	0.67		0.72
Best Parameters	{'randomforestclassifier__criterion': 'gini', 'randomforestclassifier__max_depth': 2, 'randomforestclassifier__max_features': 2}	{'logisticregression__C': 1, 'logisticregression__penalty': 'l2', 'logisticregression__solver': 'liblinear'}		{'svc__C': 1, 'svc__gamma': 100.0, 'svc__kernel': 'rbf'}






Bag of Words (BOW) Results: The best overall model for BOW was SVC but again all models were poorly performing



	Random Forests	Logistical Regression	KNN	SVC
CV (loo)	0.54 (0.50)	0.53 (0.50)		0.63 (0.48)
F1	0.76	0.72		0.80
Best Parameters	{'randomforestclassifier__criterion': 'gini', 'randomforestclassifier__max_depth': 1, 'randomforestclassifier__max_features': 1, 'tfidfvectorizer__max_df': 70, 'tfidfvectorizer__max_features': 125, 'tfidfvectorizer__min_df': 7, 'tfidfvectorizer__ngram_range': (1, 1), 'truncatedsvd__n_components': 7}	{'logisticregression__C': 0.15, 'tfidfvectorizer__max_df': 80, 'tfidfvectorizer__max_features': 150, 'tfidfvectorizer__min_df': 5, 'tfidfvectorizer__ngram_range': (1, 3), 'truncatedsvd__n_components': 7}		{'svc__C': 1, 'svc__gamma': 16.68100537200059, 'svc__kernel': 'rbf', 'tfidfvectorizer__max_df': 70, 'tfidfvectorizer__max_features': 150, 'tfidfvectorizer__min_df': 7, 'tfidfvectorizer__ngram_range': (1, 2), 'truncatedsvd__n_components': 5}



Testing on the **validation set** reveals that each of the **lexical models** **did particularly poorly as assessed by** the F1 score. However, the Gesture and Vocal models performed equally well ($F1 = 0.73$)

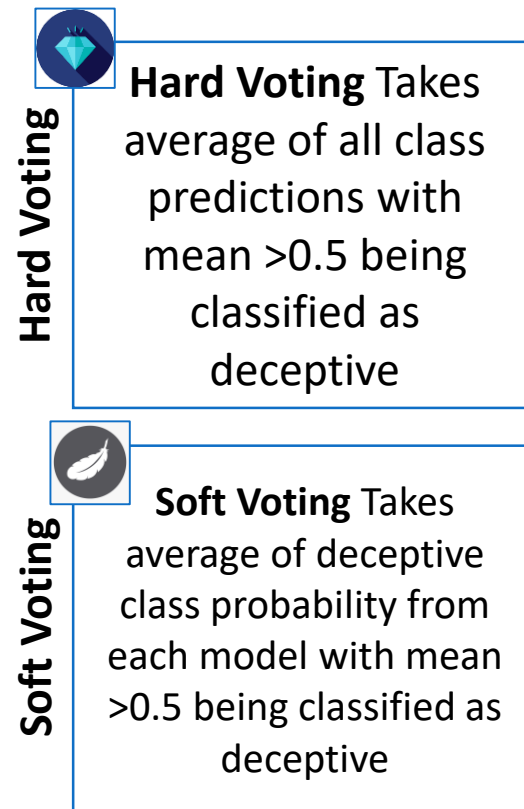
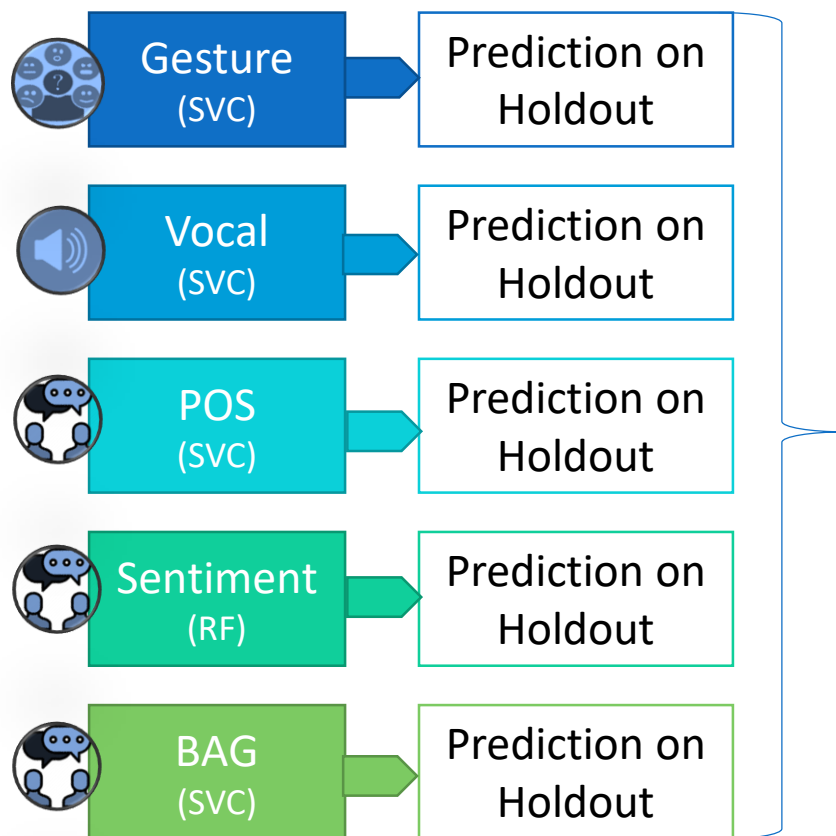
Individual Validation Model					
Modality	 <i>Gesture</i>	 <i>Vocal</i>	 <i>Lexical</i>		
			<i>POS</i>	<i>Sentiment</i>	<i>BAG</i>
<i>F1</i>	0.73	0.73	0.55	0.41	0.50



Results – Fusion Models

A blurred background image of a document or folder. The word "Results" is printed in blue, and above it is a small icon consisting of three vertical bars of increasing height, resembling a bar chart. The entire image is tilted and out of focus.

Fusion models showed high accuracy on the training data. However, **accuracy reduced significantly** on the validation set indicating all or some of the individual models overfitted



	Training	Testing
F1	0.99	0.55

Large declines for both fusion models

	Training	Testing
F1	1.00	0.73

Accuracy similar to gesture and vocal model suggesting that fusion model not improving accuracy



Conclusions



Conclusion

Conclusions

- Difficult to draw firm conclusions due to nature of data set (small; noisy; multiple contexts)
- Manual coding of gesture appears important in predicting deceptive behaviour, with particular clusters of behaviour potentially a good indicator of deception in a legal setting
- Importance of model regularisation before proceeding to any fusion
- Combination of individually poor models into a fusion model does not necessarily improve predictive performance (i.e. compensate for each other) with inspection of ROC curves vital to identifying whether models are providing any more predictive power





Future Directions



Future Directions

- **Sooner**

- Reassess individual models and apply regularisation techniques to improve generalisation of model
- Cross validate model on testing set (currently only applied on holdout)
- Assess impact of different cut-off levels on accuracy

- **Later**

- Automation of feature capture
 - Non-verbal behaviour using OpenFace
 - Normalisation of audio features to improve feature extraction
- Develop utterance level model (time period based), which will facilitate real-time detection and other feature engineering possibilities



Thank you

An illustration of a hand holding a sign. The hand is light-skinned and is wearing a grey suit sleeve with three buttons. The sign is orange and has the words "THANK YOU" written in large, white, bold, sans-serif capital letters. The background is a light blue gradient.

**THANK
YOU**

Appendix

- **Penn Treebank Project POS Tag Acronym Interpretation**

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential <i>there</i>
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun

Number	Tag	Description
19.	PRP\$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	<i>to</i>
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP\$	Possessive wh-pronoun
36.	WRB	Wh-adverb

