**Paper 1: Detecting deception through facial expressions in a dataset of videotaped**

**interviews: A comparison between human judges and machine**

**learning models.**

**1. Input kinds:**  The input for the study was a dataset of videotaped interviews of 120 participants who were asked to answer truthfully or deceptively to questions posed by an interviewer. The participants were instructed to prepare truthful and deceptive answers beforehand and were asked to provide those answers in a random order.

**2. Models and Methods used:** In accordance with the Declaration of Helsinki, the ethics committee for psychological research at the University of Padova approved the experimental procedure. The video was divided into 2 parts. The part containing free speech and the one in which subjects answer the interviewer’s questions. The “Free Speech” data set was created from the first part of the videos and the “Questions” data set from the second part. Two separate sessions (A writing session and a Videotaped Interview session) were arranged, and participants were randomly assigned to one of two videotaped Interview Conditions (Liar vs. Truth-Teller).

The first approach was a common algorithmic feature extraction technique for videos (more details follow) being fed to a linear SVM classifier, a maximum-margin linear classifier with generalization guarantees rooted in statistical learning theory . The second approach uses High-Level feature extraction using a machine learning-based tool (OpenFace). In the third

The same High level features are extracted but using an efficient Algorithm LSTM for a neural classifier suited for the classification of sequences of feature vectors. In this case, the input video is represented as a sequence of feature vectors (one for each frame). And a fully neural approach C3D, in which the network receives the raw video frames in input and the features representing the video are learned automatically from the neural network during training.

**3. Dataset**: This paper uses a dataset previously used, with a total of 62 videos of Italian volunteer participants (43 females and 19 males, aged between 20 and 29 years)

being interviewed about a past holiday. The average length of the videos was 9.56 min.

**4. Conclusion:** The study found that machine-learning models were more effective than human judges in detecting deception through facial expressions. The SVM achieved an accuracy rate of 76.3%, while the CNN achieved an accuracy rate of 68.3%. In comparison, human judges achieved an accuracy rate of 59.3%. The results of the study suggest that machine learning models have the potential to be effective tools for detecting deception in legal and forensic contexts.

**Other Insights:**

* The study found that certain facial expressions, such as eyebrow and lip movements, were more indicative of deception than others.
* The authors suggest that their study provides a valuable contribution to the field of deception detection, which has traditionally relied on human judgment.
* The study also highlights the importance of having a large and diverse dataset for machine learning models to effectively detect deception through facial expressions.

**Citation :** Monaro, M., Maldera, S., Scarpazza, C., Sartori, G., & Navarin, N. (2022). Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models. Computers in Human Behavior, 127, 107063.

**Paper 2: "Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection" published in the journal Expert Systems with Applications in May 2021.**

1. **Input kinds:** The input data for the automated deception detection system are videos of participants lying or telling the truth about a given topic. The system analyzes multiple modalities of the participants, including facial expressions, body language, and speech patterns, to detect deception.
2. **Models used**: The system uses a combination of computer vision techniques and machine learning algorithms to analyze the input data. Computer vision techniques include facial landmark detection, head pose estimation, and body posture analysis. The machine learning algorithms include support vector machines (SVM), decision trees, and random forests.
3. **Dataset**: The system is trained on a dataset of videos called the Deception In The Eyes Of Deceiver (DIED) dataset, which contains 288 videos of participants from different ethnicities and age groups. In each video, the participant is asked to answer a question truthfully or to lie about the answer. The dataset is balanced in terms of gender and contains both high-stakes and low-stakes lies.
4. **Conclusion:** The authors report an accuracy of over 80% in detecting deception using their automated system. They also analyze the contribution of different modalities to the performance of the system and find that facial expressions and speech patterns are the most informative. They conclude that their system has the potential to be used in various fields, such as law enforcement, national security, and healthcare, to improve deception detection.

**Other insights:**

* The authors also perform a cross-cultural analysis of the DIED dataset and find that there are significant differences in facial expressions and body language between cultures. They suggest that future research should include a more diverse dataset to account for these differences.
* The authors also compare their system to previous works in the field of automated deception detection and find that their system outperforms most of them. They attribute this to the use of multiple modalities and a more diverse dataset.
* The authors also discuss the ethical implications of automated deception detection and highlight the need for careful consideration of the potential biases and consequences of such systems.

**Citation:** Khan, W., Crockett, K., O'Shea, J., Hussain, A., & Khan, B. M. (2021). Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. Expert Systems with Applications, 169, 114341.

**Paper 3: Using Blink Rate to Detect Deception: A Study to Validate an Automatic Blink Detector and a New Dataset of Videos from Liars and Truth-Tellers**

1. **Input kinds**: The study used video recordings of individuals telling the truth and lying as input. These videos were recorded in a controlled environment where participants were instructed to either tell the truth or lie about a specific event that they experienced.
2. **Models used**: The study used an automatic blink detector, OpenFace 2.0, which was trained on a large dataset of videos to detect blinks accurately. The study also used statistical analysis, namely a mixed-effect model, to determine the significance of the differences in blink rate between liars and truth-tellers.
3. **Dataset:** The study collected a new dataset of 96 videos, with an equal number of videos from liars and truth-tellers, to investigate the differences in blink rate between the two groups. The videos were recorded in a controlled environment and were of high quality, which ensured that the blinks were accurately detected.
4. **Conclusion:** The study found that blink rate could be used to distinguish between liars and truth-tellers, with liars exhibiting a lower blink rate compared to truth-tellers. The study also validated the accuracy of the automatic blink detector, OpenFace 2.0, in detecting blinks accurately.

**Other insights**: The study has several implications for deception detection. Blink rate can be a useful and non-intrusive indicator of deception, which can be integrated with other cues, such as facial expressions and speech, to improve deception detection accuracy. The study also highlights the importance of using high-quality video recordings and accurate blink detection software to ensure the reliability of the findings. However, the study also notes that further research is needed to investigate the generalizability of the findings to real-world settings, where individuals may be under different levels of stress or have different motivations to deceive.

**Citation:** Monaro, M., Capuozzo, P., Ragucci, F., Maffei, A., Curci, A., Scarpazza, C., Angrilli, A., & Sartori, G. (2020). Using Blink Rate to Detect Deception: A Study to Validate an Automatic Blink Detector and a New Dataset of Videos from Liars and Truth-Tellers. In C. Stephanidis & M. Antona (Eds.), HCI International 2020 – Posters (pp. 1-6). Springer.

***Levan Sulimanov’s input:***

**Paper 4: Deception Detection in Videos**

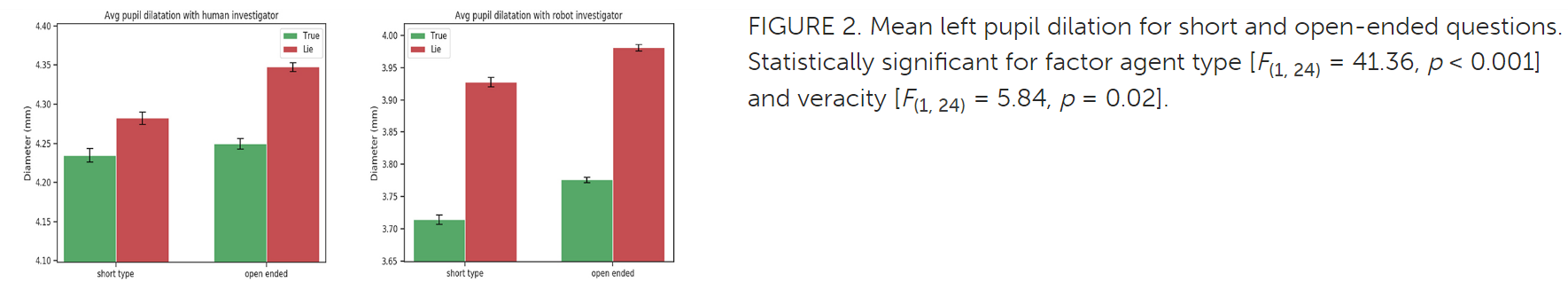
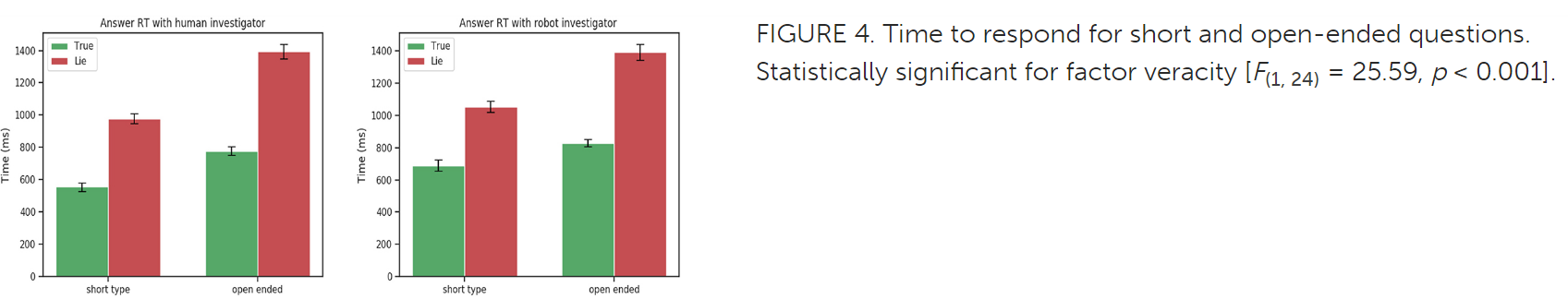
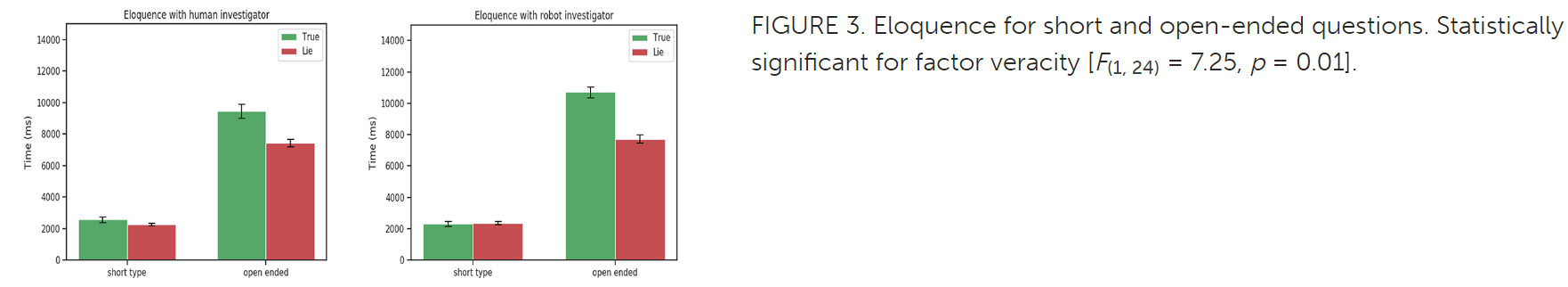
1. **Input kinds**: Primary focus is on facial micro-expressions that are thought to reflect emotions that subjects might want to hide (e.g. :Eye Brow raise).
2. **Input** is focused on multi-modal feature extraction, feature encoding, and classification. Initial feature extraction contains a sequence of video frames, corresponding audio for those frames, and audio transcription.
3. **Video input type:** The face in the data may not always be facing front or center. They employ IDT (Improved Dense Trajectory) features Using RANSAC they estimate camera motion and proceed with canceling motion in the video. **Audio Input type:** they use MFCC (Mel-frequency Cepstral Coefficient) features as their audio features - widely used for Automatic Speech Recognition (ASR). Fuse of **Video+Audio:** finally, they merge Video and Audio features into one using Fisher Vector encoding - to aggregate a variable number of features to a fixed-length vector. Transcript had little to no performance boost, so we will probably omit that part.
4. **Models used**: Binary Classifiers: (1) SVM using LibSVM, (2) Linear SVM, Kernel SVM, Naive Bayes, Decision Trees, Random Forests, Logistic Regression, and Adaboost. Polynomial Kernel for Kernel SVM showed good performance. For Naive Bayes binary classifier - normal distribution was used (feature dimensions with zero variance were removed before fitting). Logistic Regression - Binomial Distribution was used. Random Forest contained 50 trees. Adaboost used decision trees as the weak learners. All experiments used 10-fold cross validation across different feature sets and classifiers.
5. **Dataset:** 121 courtroom trial video clips. Viewing angle differs, as does the video quality and background noise. 104 videos (50 truth & 54 lies) used for training. Data contains 58 total identities.
6. **Conclusion:** The study showed that researchers were able to develop a model with
   1. Video Classification Only: model performed better by .2 AUC than human labeling (0.80 AUC).
   2. Audio Classification Only: model outperformed people by ~0.04 (AUC = 0.82).
   3. Transcript Only: ~0.1
   4. Video+Audio: 0.06 AUC performance better than humans (AUC = 0.88)

**Other insights:** Perform face alignment, in order to have each face at the same level (in order to have easier detection of micro-expressions).   
**Paper Link:** <https://arxiv.org/pdf/1712.04415.pdf>

Citation Wu, Z., Singh, B., Davis, L. S., & Subrahmanian, V. S. (2015). Deception Detection in Videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3570-3578)

**Paper 5: Deception Detection in Videos**

1. **Input kinds**: **Video**: centered on eye dilation. **Audio** - speech features.
2. **Models used**: Random Forest with cross-validation. ANOVA Analysis to check for statistical differences.
3. **Dataset:** 510 lies and 504 true statements. 20 questions for label and interviewers. People watched crime scene and then were assigned roles (to whether lie or tell the truth).
4. **Conclusion:** “The best model trained with the behavioral features achieved an accuracy of 69% and an AUCROC score of 0.74. Looking at the misclassification errors, it can be seen that the model is sensitive detecting 82% lies correctly but with a precision of 65%. Surprisingly adding the psychological traits decreased the performance of the model with a drop of accuracy and precision but a marginal gain in sensitivity.”

**Other insights: (1)** Pretty cool dilation differences when person is lying and when person is telling the truth:  
  
**(2)** Time to respond matters:  
**(3)** *Short answers VS. long answers difference - significant difference in long replies cases*:  
  
**Paper Link:** [**https://www.frontiersin.org/articles/10.3389/frobt.2019.00064/full**](https://www.frontiersin.org/articles/10.3389/frobt.2019.00064/full)

**Citation :** Gonzalez-Billandon, J., Aroyo, A. M., Tonelli, A., Pasquali, D., Sciutti, A., Gori, M., Sandini, G., & Rea, F. (2021). Can a Robot Catch You Lying? A Machine Learning System to Detect Lies During Interactions. Frontiers in Robotics and AI, 8, 694540.

**Paper 6:** Machine Learning-based Lie Detector applied to a Novel Annotated Game Dataset

1. **Input kinds**: Do not use a micro-expression approach. Instead, they base their work on facial patterns contains on single images and feed that into ML algos and then to final classifier model.
2. **Models used**: Transfer Learning by means of the embedded features from VGG-Very-Deep-16 CNN. They feeded the cropped facial to the pre-trained VGG model, and stored embedded features. These features were then used to train classifiers (SVM, Adaboost, and LDA). VGG in this case was used as the initial facial feature extractor.
3. **Dataset:** Video data collected by playing a game where convincing lie produces advantage to the one who lied. Due to game rules, researchers were able to know when person was lying or not. 19 participants. 15566 total frames: 6476 (41.6%) are lies and 9090 (58.4%) are truth. These frames correspond to a total of 417 statements, 170 of which are lies (40.8%) and 247 are true (59.2%). Hence on average, each lie statement has 38 frames, and true statements consisting about 37 frames
4. **Conclusion:** Bad results. 57% for Accuracy and 58% for F1 score. Since it is binary classification - lie or not lie, it means that random classification accuracy can be at 50%.

**Other insights:** Single image most probably will have a poor description of a lie. We need a sequence of frames.

**Paper Link:** <https://arxiv.org/pdf/2104.12345.pdf>

**Citation :** Rodriguez Diaz, N., Aspandi, D., Sukno, F., & Binefa, X. (n.d.). Machine Learning-based Lie Detector applied to a Novel Annotated Game Dataset

**Paper 7:**

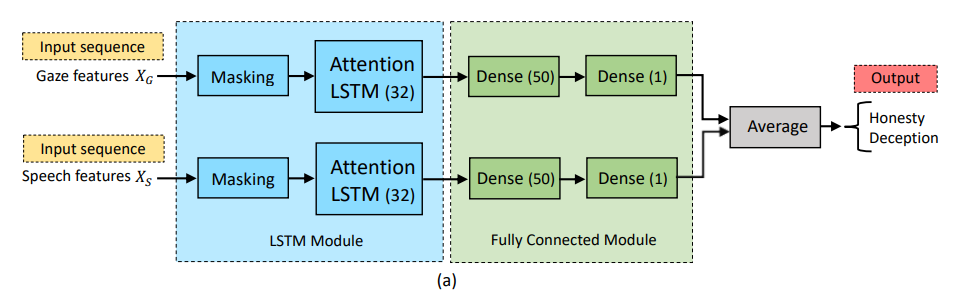
1. **Input kinds**: Video: video frames from Bag Of Lies. Audio - audio data from Bag of Lies. They set two different video length inputs: “**segment level** that referred to each of the segments of 3 s length and **turns level** that corresponded to the full recordings.”
2. **Models used**: LSTM and SVM. LSTM was used as a Feature extractor since the output from LSTM was then used as input to Dense(50 layers) -> Dense(1 as binary output).

They had 2 setups:

1. Gaze Features OR Speech features => Masking (makes sure the dummy values of the padded sequences were not used in further computations) => LSTM => Dense(50 - classifier) => Dense(1 binary output) => Truth / Lie.

2. Gaze Features => Masking => LSTM => Dense(50) => Dense(1) => Average => Lie?

/\   
 |

Speech Features => Masking=> LSTM=>Dense(50) => Dense(1)=> \_\_\_\_|  
(*which means we* ***AVERAGE Gaze*** *and* ***Speech*** *features) from each Dense(1) output.  
*

For model flow description:

“For the reference system, two strategies for combining gaze and speech modalities into an SVM framework are developed. In the first case, Late Fusion, the combination is done at the decision level. More precisely, the scores produced by each of the single-modal SVM systems are transformed to probabilities by means of a softmax operation and then averaged for producing the final score. In the second case, Early Fusion, both modalities are fused at the feature level by concatenating the corresponding compact representations into a single feature vector. ”

“For the LSTM-based systems, we also explored two combination techniques for the fusion of gaze and speech modalities, namely, Late Fusion and Attention-Pooling (AP) Fusion. As can be observed in Figure 3a, the late fusion strategy consisted of the combination of the two single systems at decision level by averaging their outputs. In the AP fusion, as shown in Figure 3b, the outputs of the Attention LSTM layers of the individual systems were combined by using a dense layer of 100 nodes followed by a final fully-connected layer of 1 neuron and sigmoid activation.”  
***Just for our note, note for summary:*** “The LSTM module consisted of a Masking layer that makes sure the dummy values of the padded sequences were not used in further computations (see Sections 2.2.1 and 2.2.2) and an attention LSTM layer with 32 hidden units and a dropout of 0.25 in order to avoid overfitting in the training process. As well, the attention parameter vector u had a dimension of 32. The Fully Connected module was composed of a dense layer of 50 neurons, and final dense layer with 1 node that was activated by a sigmoid function for performing the binary classification.”

1. **Dataset:** Bag of Lies dataset. Only used the gaze and audio (speech) modalities. Gaze data is sampled at 60 Hz and follows the GazePoint Open Gaze API specification, providing 26 channels that encode different information related to eye tracking. Audio was extracted from the video recordings by using the FFmpeg software and downsampled to 16 KHz. Some audio files contained speech from judge and suspect, so audio was trimmed to exclude initial speech and keep only suspect’s audio input. e Gazepoint GP3 Eye Tracker was used to track pupils.
2. **Conclusion:** LSTM outperformed SVM for accuracy by ~8%. AUC just by .01. Segment Level VS. Turn Level showed slight differences in performance differences: 2-3% higher for full recordings inputs in comparison to 3s inputs. But 3 second inputs are more realistic than waiting for the whole recording in real case scenarios.  
   “Secondly, focusing on LSTM-based systems, it can be observed that at segment level, gaze features were more able to correctly detect deception than honesty. In contrast, the speech modality performed considerably better for detecting truths than lies. The performance at turn level presented the same trends.   
   Finally, regarding the general comparison between the modalities in terms of AUC in the LSTM framework, gaze outperformed speech, especially at turn level, where this first modality achieved a 22.83% relative increase in AUC with respect to the second one.”

**Other insights:**  
**Paper Link:** [**https://www.mdpi.com/2076-3417/11/14/6393**](https://www.mdpi.com/2076-3417/11/14/6393)

**Citation :** Gallardo-Antolín, A., & Montero, J. M. (2021). Detecting Deception from Gaze and Speech Using a Multimodal Attention LSTM-Based Framework. *Applied Sciences*, *11*(14), 6393. MDPI AG.

**— Hinal Desai’s Inputs**

**Paper 8: Catching a Liar Through Facial Expression of Fear**

**Input Kinds:** The input to this analysis were video clippings taken from the game show ‘The Moment of Truth’. All of the video clips were recorded in a high-stake game show. Prior to the show, the contestants took a polygraph exam when they answered 50 questions. During the show, 21 of the same questions were asked again and the contestants were required to answer them in front of the studio audience. If the contestant gave the same answer to a question as they did in the polygraph exam, they moved on to the next question; lying (as determined by the polygraph) or refusing to answer a question ends the game.

**Method used:** The video clippings were imported to OpenFace. OpenFace is a Python and Torch implementation of face recognition with deep neural networks. This software automatically detects the face, localizes the facial landmark, outputs the coordination of the landmarks, and recognizes the facial Action Units(AUs). OpenFace is able to identify 18 AUs. When a person lies, the emotions that are depicted are fear and guilt, hence the focus of the study was on AUs of fear, i.e., AU1, AU2, AU4, AU5, AU20, AU26. The presence (0 or 1) and intensity (any number between 0 and 5) for each AU from OpenFace for each frame of videos was noted. MATLAB was used to determine the AUs of the emotional facial expression of fear after obtaining the AU data from OpenFace.Data was resampled using SMOTE. SMOTE is an over-sampling technique that solves class imbalance problem by using interpolation to increase the number of instances in the minority class. Resampling was necessary because the data are unbalanced, with the video clips of truth much longer than those of deception, 50,097 frames vs. 3,689 frames. WEKA, a machine learning software was used to classify the videos into truth and lie groups. The classifiers used were Random Forest, KNN, Bagging.

**Dataset:** The dataset consisted of 32 video clippings 16 of which were true and 16 lies. These video clippings belonged to 16 individuals. There were 8 males and 8 females. The video clips consist of the moments when the individuals were answering the questions, that is, from the end of the questioning to the end of the answering. The duration of the video clips ranges from 3 s to 280 s, with an average duration of 56.6 s. In total, there were 50,097 frames for truth-telling video clips and 3,689 frames for lying video clips.

**Conclusion:**



This clearly shows that Random forest outperformed both bagging and knn in terms of accuracy, precision and recall value followed by bagging and knn.

**Paper Link:** <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.675097/full>

**Citation:** Shen, X., Fan, G., Niu, C., & Chen, Z. (2018). Catching a Liar Through Facial Expression of Fear. Frontiers in Psychology, 9, 1210

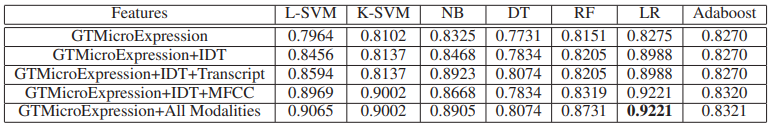
**Paper 9: Deception Detection in Videos**

**Input Kinds:** Video clippings from the original courtroom trials were used for this study. Both audio and video were used.

**Method:** Various discriminative and non-discriminative models like SVM, Naive Bayes, Decision Tree, Random Forest, Logistic Regression, and Adaboost were used to classify whether a person is lying or saying the truth.

**Dataset:** A dataset of 121 video clips out of which only 104 videos were used which had 50 true and 54 deceptive videos. The pruned videos had some manual video editing or scene change and hence were discarded. To avoid the problem of overfitting, 10-fold cross-validation is applied to the identities.

**Conclusion:** Due to their discriminative nature, SVM and Random Forest perform better than other classifiers like Naive Bayes and Logistic Regression. Different classifiers are adept at leveraging various feature modalities, which is an intriguing observation. For instance, we see that Kernel SVM performs best on MFCC features whereas Random Forest and Linear SVM both perform best on high-level microexpression features. Yet, the performance of several classifiers converges when we combine multi-modal features utilizing late fusion.

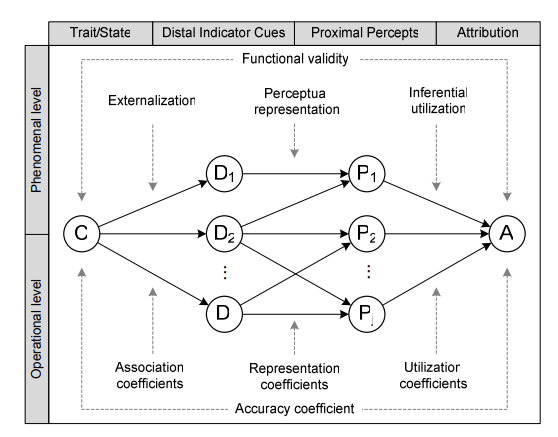
****

**Paper Link:** <https://ojs.aaai.org/index.php/AAAI/article/view/11502/11361>

**Paper 10: Video-Based Deception Detection**

**Input Kinds:** Input was video clippings that were collected from an experiment conducted at a university.

**Method:** Brunswikian Lens Model was used for deception detection.



**Dataset:** A mock theft experiment was conducted at a university with undergraduate students. A wallet was kept at the table. A few students stole the wallet, whereas a few were just present there and might not be aware of what is happening. Post this, interviews were conducted and recorded. There were a total of 40 responses that were recorded, out of which 17 participants were saying the truth and 23 were deceptive.

**Conclusion:** The experiment proved that lens models prove to be a paradigm shift in detecting truth and lies and provide a new angle for doing the same.

**Paper Link:**

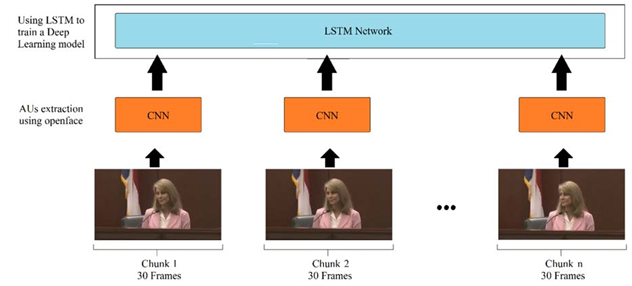
<https://www.researchgate.net/publication/226367140_Video-Based_Deception_Detection>

Citation: Jensen, M. L., Meservy, T. O., Burgoon, J. K., & Nunamaker, J. F. (2008). Video-Based Deception Detection. In Intelligence and Security Informatics (pp. 425-441).

**Paper 11: Deception Detection in Videos using the Facial Action Coding System**

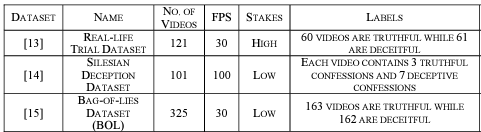
**Input Used:** Inputs from 3 different datasets were used for this study. The real-life trial dataset is a collection of real-life trial case videos downloaded from YouTube. The Silesian deception dataset consists of videos all of which were recorded at 100 frames per second. The videos neither have any audio nor the transcript for their answers provided. The Bag of Lies dataset consists of video, audio, and eye gaze.

**Method:**

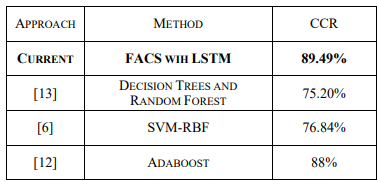


Deep learning techniques from OpenFace were used for AU detection and the output of the same was fed to LSTM for classification.

**Dataset:**

****

**Conclusion:**

****

The Facial Action Coding System worked best with LSTM.

**Paper Link:** <https://paperswithcode.com/paper/deception-detection-in-videos-using-the>

Citation: Ahmed, H. U. D., Bajwa, U. I., Zhang, F., & Anwar, M. W. (2021). Deception detection in videos using the Facial Action Coding System. Journal of Ambient Intelligence and Humanized Computing, 12(7), 6789-6800.