2021

Poker Bluff Detection Dataset Based on Facial Analysis

Umur Ciftci
Binghamton University--SUNY

Jacob Feinland
Binghamton University--SUNY

Jacob Barkovitch
Binghamton University--SUNY

Dokyu Lee
Binghamton University--SUNY

Alex Kaforey
Binghamton University--SUNY

Poker Bluff Detection Dataset Based on Facial Analysis

Jacob Feinland, Jacob Barkovitch, Dokyu Lee, Alex Kaforey, Umur Aybars Ciftci, and Lijun Yin

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

INTRODUCTION

Unstaged data with people acting naturally in real-world scenarios is essential for high-stakes deception detection (HSDD) research. Unfortunately, multiple HSDD studies involve staged scenarios in controlled settings with subjects who were told to lie. Using video footage of subjects and analyzing facial expressions instead of invasive tracking of biological processes enables the collection of real-world data.

Poker is a high-stakes game involving a deceptive strategy called bluffing and is an ideal research subject for improving HSDD techniques. Videos of professional poker tournaments online provide a convenient data source. Because proficiency in HSDD generalizes well for dissimilar high-stakes situations (unlike low-stakes deception detection), findings from poker bluff detection research will likely be applicable to other more practical HSDD applications like interrogations and customs inspections. In the hopes of encouraging additional research on real-world HSDD, we present a novel in-the-wild dataset for poker bluff detection.

DATASET

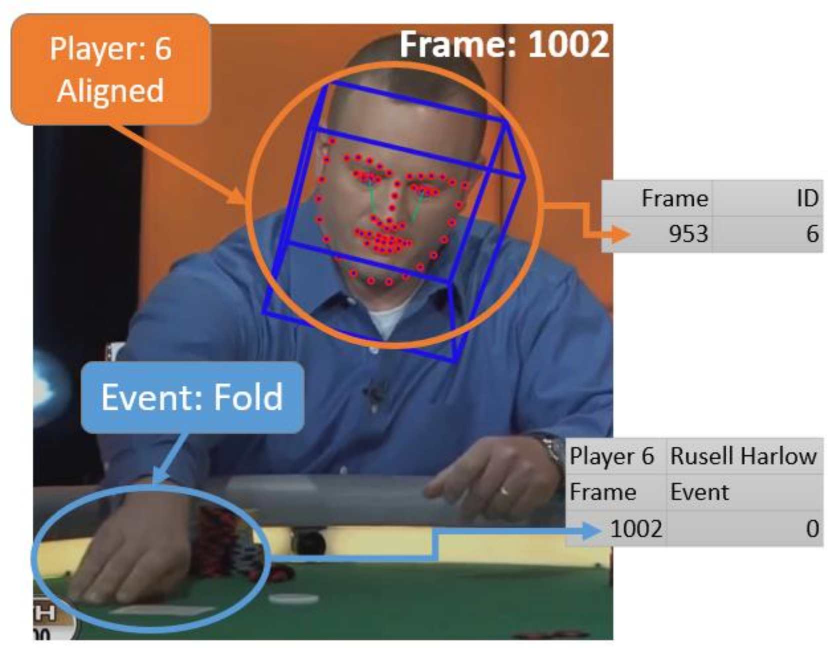

Figure 1. Bluff labeling. Whenever a player gains information, makes a bet, folds, or wins, that event is recorded with the frame number.

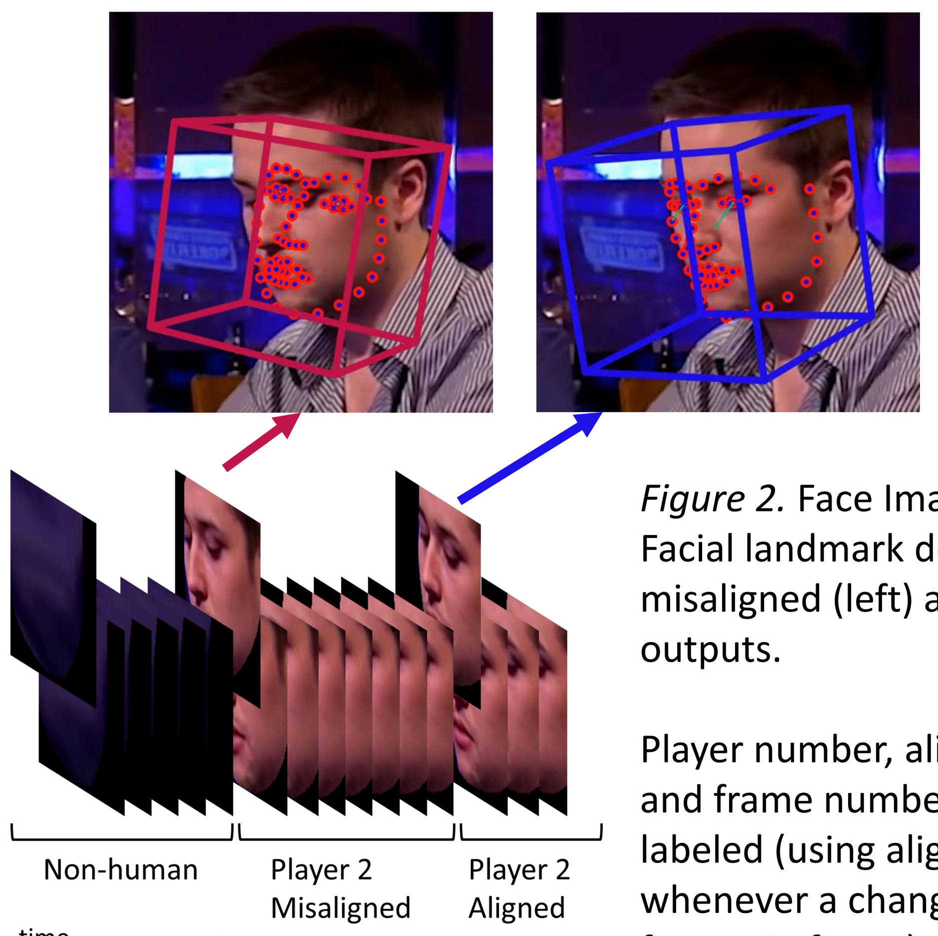

Figure 2. Face Image Labeling. Facial landmark detection [2] with misaligned (left) and aligned (right) outputs.

Player number, alignment, face ID, and frame number are manually labeled (using aligned images) whenever a change occurs (raised frames in figure).

- Four professional poker tournament videos totaling 48 min.
- Variety of head poses, lighting conditions, and occlusions

We used players' cards and bets to manually label bluffs and extracted facial expressions in over 31,000 video frames containing face images from 25 players.

DATASET

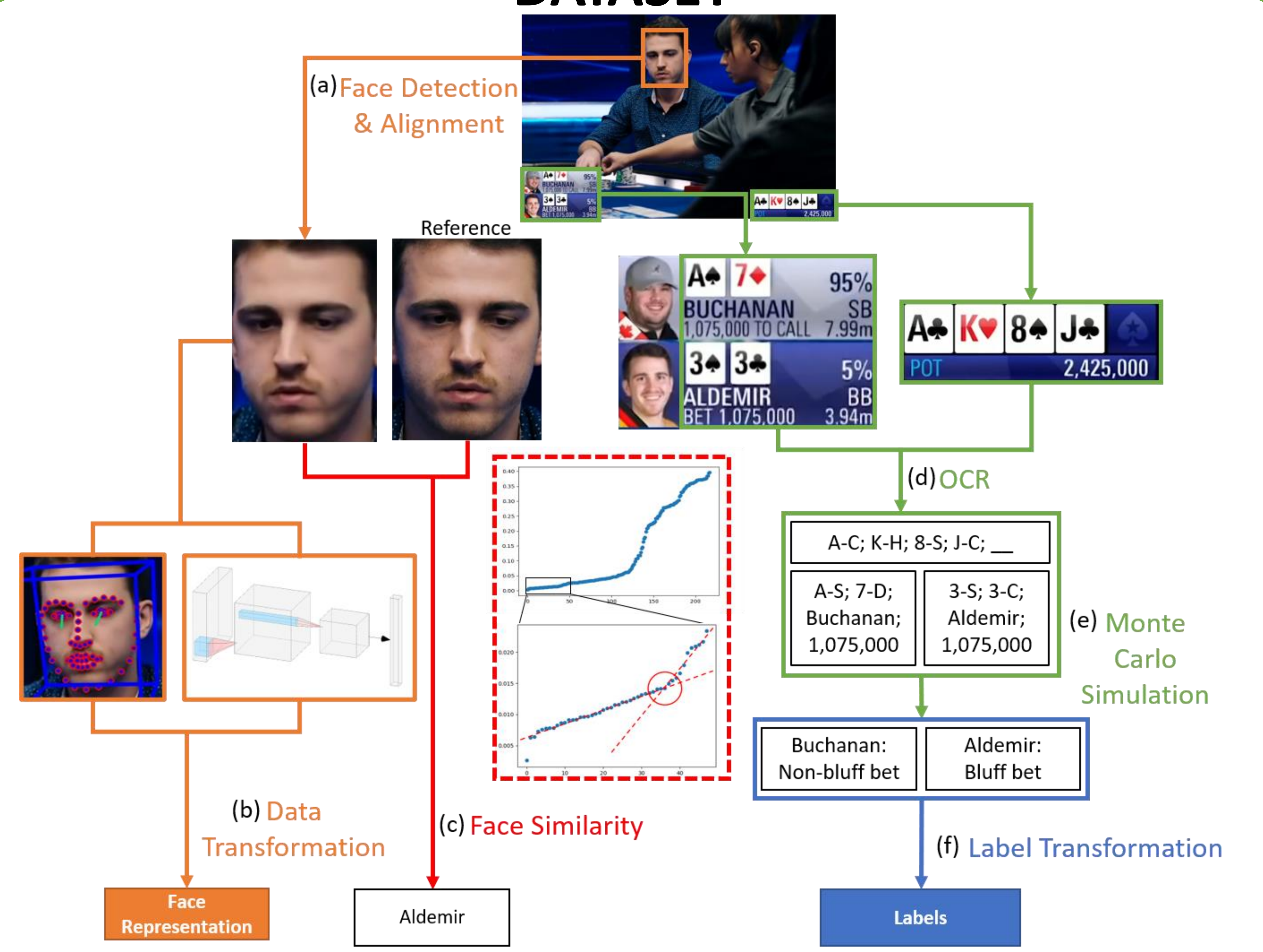

Figure 3. Labeling Process. (a) A player's face is detected and aligned. (b) OpenFace 2.0 [2] outputs AUs representing the face for the SVM model. For the CNN model, the CNN creates a feature representation from the face image. (c) We check face similarity [3] between a reference image of each player and the faces detected in the video to identify the face images by name. In the graph, many images are ranked by their similarity (left to right) to a certain reference image. An image contains the same player as the reference image if it fits within the left-most linear slope of similarity values. Values after the circled change in the graph's slope indicate a different player. (d) To retrieve on-screen data, we use Optical Character Recognition (OCR) on cropped areas of interest. (e) When there is a change in the game state (e.g. a certain player makes a bet), we categorize this event [4]. (f) We convert these event labels so that each face image has a final label to be used for image-based classification.

METHODS

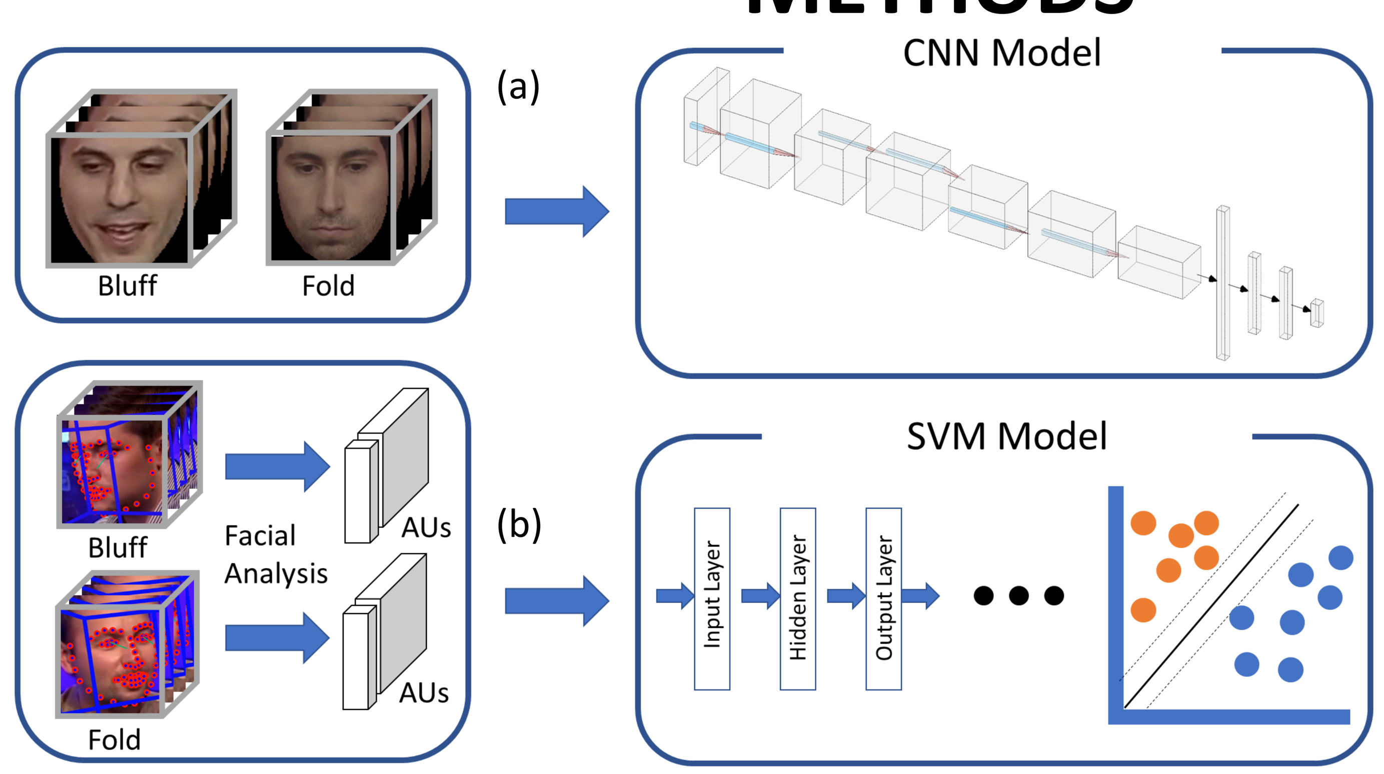

Figure 4. Bluff Detection Models. **(a)** Aligned facial images output from OpenFace 2.0 [2] are used as inputs for the CNN model trained using the bluff labels. **(b)** A feature vector of 17 AUs from OpenFace 2.0 [2] facial analysis is used as the input into the SVM model to be trained using the bluff labels.

Machine Learning Classification

1. FER Action Units (AUs) SVM vs. Aligned Facial Image CNN
2. All-Category Classification vs. Binary Bluff Classification

Images are separated into categories based on the most recent bluff event for that player. The CNN models use Leaky RELU layers, max-pooling, batch-normalization, and dropout. The binary bluff model is only an experiment and does not yield practical results since it involves heavy manipulation of data through the removal of all data from 3 out of 5 bluff categories.

RESULTS

Model	# of Bluff Categories	Accuracy
All-Category CNN	5	85.80%
Binary Bluff CNN	2	96.18%

Table 1. Final accuracy on all the models.

The Binary bluff model only uses faces categorized as clearly bluffing or clearly not bluffing. The All-category model uses these two categories as well as one for a bet with an ambiguous bluff status, one for before viewing cards, and one for before making a bet.

Overall accuracies from binary models were higher than all-label models, including training accuracies for the SVM models.

The high accuracy for the Binary bluff model suggests that bluffing can be discerned using facial analysis.

Our All-category CNN model achieves a significantly higher accuracy than [1].

Model	Accuracy
All-Category CNN	85.80%
All-Category SVM	56.23%
[1] Regular	66.81%
[1] Balanced	59.38%

Table 2. Comparison of results to [1], which predicts folds using a decision tree model. The balanced values are the accuracies after the folds were scaled to have equal weights to the calls and raises within the video. The classification rate is the accuracy of the best model.

CONCLUSION

After our promising baseline results, we believe this dataset will allow future in-the-wild bluff detection research to achieve higher deception detection rates, which will enable the development of techniques for more practical applications of HSDD such as in police interrogations and customs inspections.

Future Works:

We could extensively evaluate and modify our models to ensure that bluff classifications are independent of how frequently each bluff label occurs both overall and for each player.

We are planning to combine multiple facial modalities into a single model (e.g. face AUs and CNN features).

We could test our trained model on other scenarios of high-stakes deception detection such as videos from police interrogations or court hearings.

REFERENCES

[1] D. Vinkemeier, M. Valstar, and J. Gratch. Predicting folds in poker using action unit detectors and decision trees. In 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), pages 504–511, 2018.
[2] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), pages 59–66, 2018.
[3] S. I. Serengil and A. Ozpinar. Lightface: A hybrid deepface recognition framework. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pages 1–5, 2020.
[4] Jonathan Rubin and Ian Watson. Computer poker: A review. Artificial Intelligence, 175(5):958–987, 2011. Special Review Issue.