Recognizing Cultural Differences in Fear Social Signals

Arda Cifci
aca261@sfu.ca

Tushrima Kelshikar
tkelshik@sfu.ca

Zihan Yu
yuzihany@sfu.ca

Recognizing Differences in Fear Social Signals, Final Project, April 17

1 ABSTRACT

Modern social signal recognition technology can recognize the primal emotion fear accurately, but a majority struggle to account for the nuances that come with emotional social signal recognition. A person's culture and sub-categorical type of fear can drastically alter an individual's social signal profile, making the currently employed "one size fits all" approach to fear social signal recognition vastly oversimplified. In this project, we created a dynamic fear social signal recognition algorithm that can recognize the cultural differences and fear subcategories in individuals experiencing fear. We utilize a set of six trained neural networks classifiers that each recognize different facial features, body language, sound social signals and the cultural differences between Canadian, Turkish, Chinese, and Indian cultures associated with emotional fear. Our algorithm would be able to distinguish between the four different cultures and recognize three different subcategories of fear. Our work provides the ability to recognize and account for the nuances that come with fear social signal recognition and allows for a higher-level interpretation from dynamic fear social signals.

2 INTRODUCTION

Understanding the subtle expressions of fear across different cultures is not just an academic pursuit but a way to improve global communication and empathy. While affective computing has recognized basic emotions, the complexity of emotional expressions, particularly fear and its subcategories, remains simplified in current technologies. This simplification often results in a one-dimensional interpretation of fear that ignores cultural and contextual subtleties inherent in human interactions. [1]. For instance, what may be perceived as mild nervousness in one culture could be interpreted as profound terror in another, each accompanied by distinct social signals like facial expressions, body language, and vocal nuances [2].

The primary challenge in advancing social signal recognition technology lies in the current "one size fits all" approach, which fails to consider the rich tapestry of human emotion molded by cultural backgrounds. Research indicates significant variations in how emotions are expressed and perceived across different cultures, affecting the accuracy of emotion recognition systems [3]. However, we were not able to find any existing databases that would align with our research topic.

In response to these gaps, our project introduces a novel algorithm designed to recognize and differentiate fear's sub-categories and the associated cultural social signals among Canadian, Turkish, Chinese, and Indian populations. By leveraging a combination of six neural network classifiers, each tailored to detect unique sets of facial features, body movements, and vocal characteristics specific to these cultures, our model represents a significant leap from generic algorithms. This approach not only respects the complexity of human emotions but also caters to the global need for more culturally sensitive technologies in fields ranging from security to healthcare and customer service.

Through this research, we aim to bridge the gap between universal emotion recognition models and culturally specific emotional intelligence. Our work not only adds to the growing body of knowledge in cross-cultural psychology and affective computing but also sets a precedent for future research in creating more empathetic and understanding human-machine interactions. By acknowledging and addressing the cultural nuances in emotional expression, our algorithm stands to revolutionize the field of social signal recognition, paving the way for more nuanced and culturally aware systems.

3 APPROACH

The video samples constituting our dataset were extracted from YouTube using Open Broadcast Software (OBS) to record, targeting a collection of natural and staged expressions of fear to diversify the training data for our classifier. We manually selected videos where the individuals displayed fear, ensuring a mix of naturally occurring fear and its sub categories for the purpose of the video. This selection process was guided by predefined criteria focusing on the authenticity and clarity of fear expression.

Each video was annotated by a team of three annotators hailing from the represented demographics (Canadian, Indian, Chinese, and Turkish). This demographic diversity in our annotators was crucial to minimize cultural bias in identifying fear expressions. Annotators annotated based on the individual's apparent nationality and the observed fear response. The videos classified fear into three subcategories: nervousness, threatened, and surprised. Discrepancies were resolved through a consensus meeting, ensuring high inter-annotator agreement.

Pre-processing involved normalizing video lengths to a consistent format of 1-10 seconds to streamline computational processing and feature extraction. Videos were converted into a frame-by-frame analysis where each frame was subjected to the following processing steps:

3.1 Audio/Sound Extraction

Audio tracks were separated from each video sample and analyzed using 'librosa'. We extracted key features such as mean amplitude, dominant frequency, pitch, and Mel spectrogram.These were calculated to capture the variability in sound expression, which can be indicative of different fear responses.

3.2 Body Movement Extraction

Using the OpenPose software, which implements the BODY25 model, we extracted coordinates for 25 body joints. These coordinates were used to calculate the movement dynamics during the fear response, providing quantitative data on how fear affects body posture and movement. Principal Component Analysis (PCA) was then applied using the PCA module from sklearn.decomposition to reduce the dimensionality of the feature set, enhancing the training efficiency and effectiveness of the classifier.

3.3 Facial Feature Extraction

Facial analysis was conducted using an advanced facial recognition software, OpenFace that detects gaze direction, eye locations, and facial action units (AUs). We averaged these features over the video duration to reduce data dimensionality and focus on the most prominent facial expressions of fear.

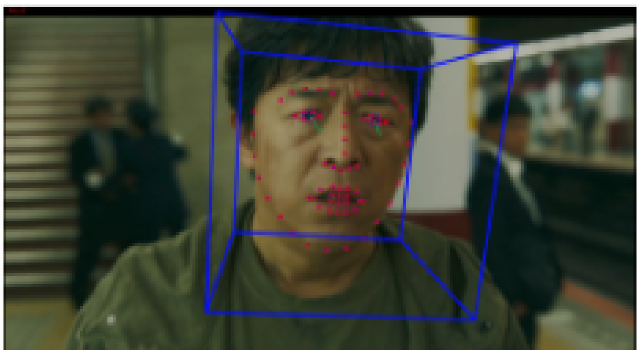

Figure 1: Example of facial data collected via OpenFace

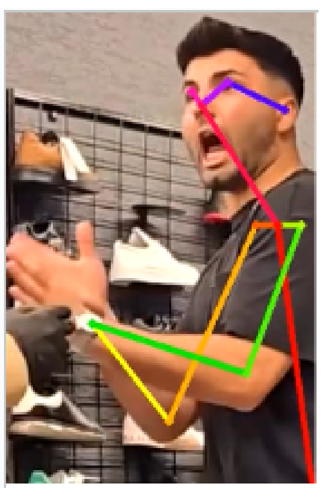

Figure 2: Example of body data collected via OpenPose

We utilized MLPClassifier from sklearn's neural network library for classification due to its ability to learn non-linear models and handle high-dimensional data effectively. We used the following line of code, mlp = MLPClassifier(hidden_layer_sizes=(100,), max_iter=300, activation='relu', solver='adam', random_state=1), which has the following parameters: 'hidden_layer_sizes': This parameter specifies the architecture of the hidden layers in the MLP network. The tuple (100,) indicates that there is one hidden layer consisting of 100 neurons. A single layer with 100 neurons is a common choice for a moderately complex task. 'max_iter': Here, it is set to 300, meaning the training process will run for a maximum of 300 epochs before stopping. This helps in fine-tuning the model by providing ample opportunity for the learning algorithm to converge on an optimal solution. 'activation': The parameter used for the hidden layers, relu, which stands for Rectified Linear Unit, is a popular choice due to its efficiency and effectiveness.

Extensive use of Python libraries such as numpy, pandas, matplotlib, seaborn, and sklearn ensured robust data handling, processing, and visualization. These tools were integral in managing high-dimensional data, implementing machine learning algorithms, and visualizing results effectively.

The core of our approach is a multi-modal machine learning classifier designed to integrate audio, body movement, and facial features to predict the culture as well as fear subcategory exhibited by individuals in the videos. We utilized a deep neural network architecture due to its proficiency in handling high-dimensional data and its capacity for feature learning.

4 DATASET

Our dataset contains 200 dynamic video samples. Each video sample is short, between one and ten seconds long, each demonstrating an individual experiencing fear. Our video samples were all collected from the video sharing website, Youtube, and are a mix of natural expressions and staged expressions of fear. The dataset uses natural and staged samples in order to prepare and train our algorithm's neural networks for both types of fear expression recognition in case the algorithm encounters them. Of the 200 fear video samples, 50 samples are from Canadian individuals experiencing fear, 50 samples are from Indian individuals, 50 samples are from Chinese individuals, and 50 samples are from Turkish individuals. We collected video fear samples from these countries because each one is from a different geographical world region and all have unique cultures which would adequately encompass any cultural differences in fear social signals.

Additionally, the dataset video samples are also split between three fear subcategories, nervousness, threatened, and surprised. We collected these specific subcategories because it encapsulates a wide range of social signals that can come from fear. For example, with nervousness we see more reduced movement, with threatened we see more defensive behavior, and with surprise we see increased movement and sounds. Furthermore, when any individual experiences fear these three subcategories are some of the most common forms of fear expressed. Therefore, being able to recognize these fear subcategories would be beneficial.

4.1 Features

Our dataset has three distinct sets of features. Audio/sound features, Body movement features, and facial features. From the audio/sound of the video samples we collected 8 features from each video. The mean values of the amplitude, dominant frequency, pitch, and mel spectrogram. The standard deviation of the amplitude and mel spectrogram. Lastly, the individual in the video's country and fear subcategory experienced. From each individual's body movements in the samples we collected xy coordinates for

25 different body parts following OpenPose's BODY25 model. This would include arms, legs, feet, head, shoulders, joints, chest, etc. Additionally, we collected the individual's country and fear subcategory experienced. From each individual's face we collected a multitude of features including gaze, eye locations, and Action Units (AU's). Due to the sheer size of the data from the videos, we took the mean of the features to store a single value instead of tens of thousands.

4.2 Analysis

From our analysis of the dataset, we found that Chinese individuals expressed nervousness, threatened, and surprised subcategorical fear mainly through their facial expressions. Neural networks achieved 71% accuracy in classifying Chinese individuals' fear sounds, compared to 70% for Turkey, 25% for Canada, and just 18% for India. Additionally, we found that Turkish individuals expressed fear and its subcategories with more body movement and vocals, less facial expressions. Neural networks recognized Turkish facial fear with 50% accuracy, Chinese with 33%, Canadian with 20%, and lowest in India at 12%.

Furthermore, we saw that Indian individuals express fear more moderately across all features. They were neither loud or quiet, had small movement or large movement, or showed extensive facial expressions in comparison with the other cultures. With neural networks body movement classification, we saw China and Turkey had a 80% accuracy rate each, India with 54% and Canada with 25%.

Canadian individuals had a mix of everything. As Canada is a melting pot of culture and people, each Canadian video of the dataset had diverse features. Some video samples had a lot of body movement, others had extensive facial expressions, and some had neither. Overall, the Canadian video samples were the most diverse and distinct from the rest of the video samples and the accuracy observed from our classifiers reflect that.

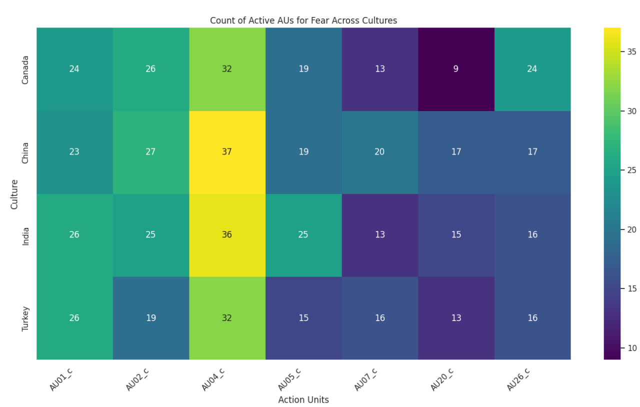

Figure 3: Heat map of AU activation between cultures

5 EXPERIMENTS AND RESULTS

5.1 Hypotheses

Our project is predicated on two main hypotheses:

Cultural Sensitivity: A neural network classifier trained with culturally specific data can better identify the nuances of fear in different cultural contexts compared to a generalized model.
Modality Integration: Integrating multiple modalities (audio, and visual) will improve the accuracy of emotion recognition systems, particularly for subcategories of fear.

5.2 Model Architecture and Training

We utilized 'MLPClassifier' from sklearn's neural network library, configured with one hidden layer of 100 neurons, using the 'relu' activation function and 'adam' solver.Models were trained separately for predicting fear subcategories and the country of origin, using a 80%-20% training-test split. We applied 20-fold cross-validation to estimate model performance and ensure generalizability. Each model was trained over 300 epochs.

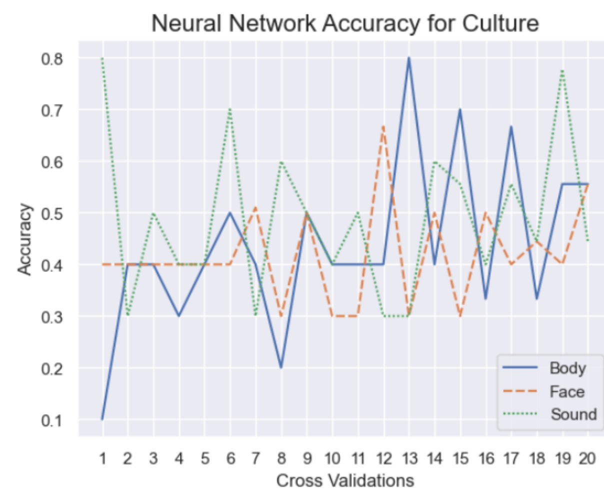

Figure 4: Cultural model cross-validation accuracy

5.3 Results

Our designed models pay close attention to the different ways different cultures express subcategories of fear. They are more accurate, which suggest that they are good at recognizing different subcategories of fear from different cultures. The accuracy averaged at 50% across the spectrum of cultures included, highlighting the capability of these models to process and understand the complex layers of cultural influence on emotional expression. The reliability of these culturally attuned models was further confirmed through an 20-fold cross-validation procedure. This procedure ensured a thorough vetting of the models' effectiveness, as each data segment underwent both training and testing. The outcome was a consistent set of accuracy results across the numerous folds, reinforcing the models' stability and reliability. This uniformity in cross-validation outcomes is significant. We hit accuracy marks like 45%, 48%, and even reaching up to 53% in some folds. It implies a well-rounded model not overly fitted to a specific dataset. The consistent results suggest that the models have a good base for understanding cultural differences in expressing subcategories of fear, hinting at their potential usefulness. There's space for improvement, but it's a good start for applications that need to consider cultural differences in emotion recognition.

6 DISCUSSION

Our research provided several interesting findings regarding the recognition of fear subcategories across different cultures using our social signal recognition model. One notable observation was variability in the performance of our classifiers based on cultural data, particularly with Canadian samples. The Canadian data showed a unique blend of fear expressions that often combined facial cues, body movements, and vocal characteristics. This inconsistency can be related to the "melting pot" nature of Canadian society, where a single expression of fear might combine elements from multiple cultural backgrounds. This diversity, although valuable, it introduces variability that our current model struggled to generalize effectively.

Unexpectedly, our model under-performed in recognizing fear expressions in videos where only facial data was available, such as in cases where the subject's face was significantly zoomed in, eliminating the visibility of the body. This limitation was pronounced when using OpenPose, a tool that tracks body movement. OpenPose requires visibility of various body joints to function effectively, and when these joints were not visible due to the camera angle or the frame composition, the recognition accuracy significantly dropped.

Moreover, the effectiveness of our system varied under different lighting conditions. In instances where video samples were poorly lit or the environment was too dark, both OpenPose and OpenFace struggled to accurately detect and process fear-related cues. This suggests that the systems depend heavily on good visual quality to perform optimally. The sound analysis, while robust in clear audio conditions, showed limitations in noisy environments or where the sound was not distinctly discernible from background noise. This often led to misclassifications, especially when subtle sound features critical for distinguishing between fear subcategories were masked by ambient sounds.

To enhance the effectiveness and reliability of our social signal recognition model, a few possible routes would be to expand the dataset, which would mean increasing the number of video samples per culture and fear subcategory would help improve the robustness of the model. More data would allow the neural network to better learn and generalize across a more comprehensive array of expressions and conditions. Enhanced feature extraction techniques: Improving tools like OpenPose and OpenFace to handle videos with zoomed-in faces or poor lighting conditions would be crucial. Advanced preprocessing techniques to enhance video quality or more sophisticated models capable of extracting features in sub-optimal conditions could address these issues.

7 CONCLUSION

This paper introduces our fear-subcategory and cultural dataset and social signal recognition model. Our video samples underwent three stages of processing to extract sound, body keypoint, and facial features for our dataset. We trained and utilized multi-layer-perceptron neural networks to classify our features into cultural and fear subcategory outcomes. Our analysis and classification demonstrated that recognizing different cultural fear social signals was not only possible, but that there are a plethora of features that could be used to recognize these differences. However, there are limitations in our current model's accuracy, mostly due to not acquiring enough video samples for each culture and subcategory. If we were to continue this project in the future, we would work to improve the accuracy of our social signal recognition model and look into including more cultures and fear subcategories.

8 APPENDIX

8.1 Datasheets for Datasets

3.1 - Motivation

• For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

This dataset was created for analysis and classification of the three fear subcategories nervousness, threatened,and surprised with regards to four different cultures, CanadIan, Indian, Chinese, and Turkish. We wanted to fill the empty gap in cultural and subcategorical fear datasets for classification as a vast majority don't take these two nuances into account.

• Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

Final project group 20 created this dataset from SFU's CMPT 419/724 class.

• Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

No funding was involved

3.2 - Composition

• What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

Inside report body - Dataset section

• How many instances are there in total (of each type, if appropriate)?

Inside report body - Dataset section

• Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

Since we are obtaining nervous, threatened, and surprised fear examples from Canadian, Turkish, Chinese, and Indian cultures, we are obtaining a subset sample of all fear and culture. The larger set is all subcategories of fear and all cultures.
Validation is part of the report - approach
We can't do the larger set due to the sheer size of fear examples we would need and the many different types of cultures. We would probably need ten thousand videos to truly cover the larger set.

• What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.

We included the raw video samples and a set of features for each video. Feature and raw video description can be found in the report.

• Is there a label or target associated with each instance? If so, please provide a description.

Each video instance targets an individual experiencing one of three types of fear, nervousness, threatened, and surprised. Additionally, the samples are either of Canadian, Turkish, Chinese, or Indian descent.

• Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Some features are missing from specific videos due to OpenFace or OpenPose not being able to recognize parts of the individual in the videos. For example, there could be no legs in the shot or the individual turned around so their face was no longer visible.

• Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

Any relationships between our dataset is made through the folder structures of our dataset. Canadian videos are grouped together with other Canadians, Turkish is grouped with other Turkish etc.

• Are there recommended data splits (e.g., training, development / validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

We primarily used the random seed 42 with our training and testing split, but others can be used to achieve similar results.

• Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

Regarding the sound portion of our dataset, there might be at times background music or sounds that are sources of noise. We processed them to reduce this noise as much as possible.

• Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

This dataset is self-contained and uses no external resources.

• Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.

No, there are no confidential data in our dataset.

• Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why. If the dataset does not relate to people, you may skip the remaining questions in this section.

Yes, the videos are of people experiencing different kinds of fear. This may cause anxiety in individuals viewing the dataset or cause them emotional fear.

• Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

Answered in the report body - Dataset

• Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.

No, direct individual information in the dataset was not taken except for their country of origin. It would be impossible to identify individuals in the dataset by our features and samples alone.

• Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.

This data contains country of origin information for the cultural aspect of our project. Besides this no other sensitive information was collected.

3.3 - Collection Process

• How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

Answered in the report.

• What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?

Answered in the report.

• If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

Answered in the report, we sampled based on the three fear subcategories nervousness, threatened, and surprise and the individuals associated culture. Videos were picked essentially randomly based on guidelines, but ensuring the videos were usable in our dependencies.

• Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

Our final project group 20 student members, 3 people total. No compensation as it was a school project.

• Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

The data was collected throughout the month of March 2024, between the second and third weeks. As our video samples are youtube videos, They were all uploaded during different days, months, and years.

• Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation. If the dataset does not relate to people, you may skip the remaining questions in this section.

No ethical review processes were needed.

• Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

The data was collected via a third-party website, Youtube.

• Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

No, they were not notified.

• Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

No, they did not consent to the collection and use of their data as they were not notified.

• If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or 8 Gebru et al. for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

No consent was obtained

• Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

No, an analysis was not conducted.

3.4 - Preprocessing/cleaning/labeling

• Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.

Answered in the report

• Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

Yes, inside the folder the raw videos were stored.

• Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.

Yes, the software is available inside our project folder. Any installation requirements were also mentioned in the README file.

3.5 - Uses

• Has the dataset been used for any tasks already? If so, please provide a description.

The dataset was used in our own project, Recognizing Cultural Differences in Fear Social Signals. Details are answered in the report.

• Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

No, only our submitted paper to CMPT 419/724 canvas submission.

• What (other) tasks could the dataset be used for?

Since our dataset contains facial features, body features, and sound features, there could be many uses for our dataset. Perhaps it could be used for discerning differences between cultures or used in a wider set of emotional social signal classification where there is more emotions than just fear.

• Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?
No

• Are there tasks for which the dataset should not be used? If so, please provide a description.

Please don't use this dataset to discriminate between cultures of people, I.e. don't use it to exclude Turkish people from social signal recognition, etc.

8.2 Contributions

Arda Cifci:
-Overall, handled everything that had to do with the body. This includes processing and extracting the body features and classifying them into our neural network models and analyzing them.
-Collected 50 Turkish sample videos and 18 Canadian sample videos.
-Created the body-keypoint "Body_Processing" preprocessing file.
-Processed 200 videos using OpenPose to extract body key-point data.
-Created the body features portion of the dataset.
-Worked on the body portions of classifying inside the "Fear_Classification" file.
-Did the abstract, dataset, conclusion, and datasheets for datasets questions of the report. Edited project code and the report sections that weren't done by me.

Tushrima Kelshikar:
-Collected 50 video samples featuring individuals of Indian ethnicity and 15 videos of Canadian individuals.
-Utilized the OpenFace tool to annotate the 200 collected videos, focusing on extracting facial expressions.
-Implemented classifier and trained using multi-layer perceptron (MLP) neural network model for face data, Face_features.csv.
-Created a preprocessing CSV file for the face features.
-Contributed to the report writing.

Zihan Yu:
-Collected 50 video Chinese samples videos and 17 Canadian samples videos.
-Transfer mp4 files to mp3 files.
-Using librosa Extracted the sound features from mp3 files.
-Did the Experiment and result for the report.
-Created a CSV file for sound features.

9 REFERENCES

[1] P. Ekman, "An argument for basic emotions," Cognition and Emotion, vol. 6, no. 3–4, pp. 169–200, May 1992, doi: https://doi.org/10.1080/02699939208411068.

[2] H. A. Elfenbein and N. Ambady, "On the universality and cultural specificity of emotion recognition: A meta-analysis.," Psychological Bulletin, vol. 128, no. 2, pp. 203–235, 2002, doi: https://doi.org/10.1037/0033-2909.128.2.203.

[3] D. Matsumoto, "Cultural Influences -on Facial Expressions of Emotion CROSS-CULTURAL RESEARCH ON THE EMOTIONS." Available: https://davidmatsumoto.com/content/1991%20Cultural%20Influences%20on%20Facial%20Expressions%20of%20Emotions.pdf

[4] R. Javadi, "The Many Faces of Anger: A Multicultural Video Dataset of Negative Emotions in the Wild," 2021. Accessed: Apr. 18, 2024. [Online]. Available: https://summit.sfu.ca/_flysystem/fedora/2022-12/etd21669.pdf

[5] X. Deng, S. An, and Y. You, "Cross-cultural differences in the processing of social and non-social positive emotions: An ERP study," Current Psychology, Jan. 2022, doi: https://doi.org/10.1007/ s12144-021-02604-8.

[6] Z. H. Pugh, S. Choo, J. C. Leshin, K. A. Lindquist, and C. S. Nam, "Emotion depends on context, culture and their interaction: evidence from effective connectivity," Social Cognitive and Affective

Neuroscience, vol. 17, no. 2, Jul. 2021, doi: https://doi.org/10.1093/scan/nsab092.