Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to patients and screenshots, if applicable? [Yes]
   (b) Did you describe any potential patient risks, with links to Institutional Review Board (IRB) approvals, if applicable? [Yes]
   (c) Did you include the estimated hourly wage paid to patients and the total amount spent on patient compensation? [Yes]

A Author Statement

Upon future release of the data, NIH bears all responsibility in case of violation of rights, etc., and confirmation of the data license. Wherever possible, all data are de-identified; however, data consisting of facial image frames are inherently classified as Personally Identifying Information (PII) and as such are unable to be de-identified. All data is stored under the control of an internal NIH database accessible to other NIH researchers for future research. NIH will share access to outside researchers under their discretion. As part of the informed consent process, governed by the NIH Institutional Review Board (IRB), patients have given permission for their data to be shared under these restricted conditions for future research.

B Acknowledgements

We are grateful for the support and advisement received on the development of the ISS Study Design and Protocol. We acknowledge Mason Rule and Susan Wroblewski at the National Institutes of Health, and Lauren Neal, Jeremy Walsh, and John Larson at Booz Allen Hamilton.

C Ethical Considerations

The generation of this dataset has brought multiple ethical concerns to our attention. This section is intended to share our observation of ethical concerns to raise awareness of the issues and stimulate conversation around potential mitigation. We will discuss the following ethical concerns:

- Use of Fitzpatrick Skin Scale
- Data collection biases
- Video lighting
- Camera quality and availability
- Facial Recognition Algorithm Bias and Surveillance

Use of Fitzpatrick Skin Scale The Fitzpatrick Skin Scale was originally developed for dermatological use to measure the sensitivity of skin burn during phototherapy and has recently been critiqued for its conflation with race and ethnicity [59]. Further, the scale has been found to overestimate the prevalence of Type IV skin classification in African Americans [60]. In a sample of 2,086 California Black adults, 59% categorized themselves as not applicable to any of the four skin types and only 26.8% identified as Type IV, the darkest skin type. The authors assert that the Fitzpatrick Skin Scale is more in line with Caucasian experiences of sun-reactivity and not Asians, Arabs, and African Americans. The visual grouping of patients into “Dark" (Skin Types IV - VI) or “Light" (Skin Types I - III) may be too restrictive and biased in terms of broadening our diversity of patients.

Data Collection Biases As we collected videos from patient owned devices (e.g. mobile phones, tablets, etc), we immediately noticed wide ranges of video characteristics. Videos were captured using various illuminations (e.g., front lighting vs. backlighting), varying instances of whether a patient is (or is not) wearing glasses, and masks are occasionally worn in videos due to the global Covid-19 pandemic. We also noticed a variety of video quality and resolution which is most attributed to the condition of the physical device hardware used to capture the video. As we continue to use this data to train models, it may be useful to quantify the extent of these many video characteristics and store them as metadata. Having the video characteristics stored as metadata would allow for bias detection techniques to be for identification and (if possible) mitigation of data biases. Although an analysis of socioeconomic backgrounds of each patient has not been conducted, it is fathomable that those in lower socioeconomic groups may not have access to the latest smartphones and devices. This may lead to lower video quality and as a result, model results that are not representative across the broad spectrum of patients. Intrinsic penalization based on video quality conflated with socioeconomic status should be avoided.

Facial Recognition Algorithm Bias and Surveillance Existing bias is always a concern when using pretrained models for either direct inference or transfer learning. Facial recognition algorithms are no different. Researchers [65] explore issues with facial recognition algorithms explored through the context of the (already concerning) Fitzpatrick Skin Scale. Further, the research towards chronic cancer facial pain detection is not intended for surveillance purposes. A potential misuse would be for monetary purposes of estimating medical risk and liability by analyzing and estimating pain without consent.

D Preprocessing

D.1 Corrupt Data

Patient 0009’s videos were unreadable and could not be processed. This consists of six potential video submissions. Patient 0015 submitted four videos, but never submitted pain scores at the start of the submission.

D.2.5 Model Transforms

We use PyTorch V1.6.0 for implementing transforms. Limited transforms were conducted to only include `RandomResizedCrop()` and `RandomHorizontalFlip`. All images for the facial images and the Mel spectrogram were resized to 224 x 224. Face crops were originally 160 x 160 following MTCNN detection on frames. Images were normalized to ([0.485, 0.456, 0.406], [0.229, 0.224, 0.225]). No image augmentation or synthetic images were used in training.

E Additional Data Analysis

In Table 6 we provide counts for the total number of videos, video frames, seconds of video, average seconds per video, and 4-sec. chunk `.wav` files for each of the 29 patients. Counts are also provided by the original four levels of pain ("None", "Low", "Moderate", "Severe"), in addition to the two levels ("No Pain", "Pain") used for the binary classification models.

Table 6: **Descriptive Statistics for ISS Data Across 29 Enrolled Patients**. The number of Pain Score submissions per patient is equivalent to the number of Videos, since pain scores are submitted at the start of each submission.

Patient	Videos	Total Frames	Total Seconds	Avg Sec/Video	4 Sec. Chunks	(4-Pain) Level	(2-Pain) Level	Sex	Skin Tone
1	23	1674	167.23	7.27	53	None	No Pain	M	IV - VI
2	23	1721	172.24	7.49	54	None	No Pain	M	I - III
3	5	2558	255.91	51.18	67	None	No Pain	M	I - III
4	5	1605	160.08	32.02	43	Moderate	Pain	M	I - III
5	5	2552	255.17	51.03	67	Moderate	Pain	M	I - III
6	1	193	19.39	19.39	5	Severe	Pain	F	IV - VI
7	26	5398	539.61	20.75	148	Low	Pain	M	IV - VI
9	n/a	n/a	n/a	n/a	n/a	Low	Pain	F	I - III
8	25	6938	694.21	27.77	185	Moderate	Pain	F	I - III
10	18	3523	352.43	19.58	95	Severe	Pain	F	IV - VI
11	26	35249	3525.39	135.59	896	Severe	Pain	M	I - III
12	55	41373	4136.35	75.21	1064	None	No Pain	M	I - III
13	28	6953	694.98	24.82	187	None	No Pain	M	I - III
14	43	16685	1669.08	38.82	435	None	No Pain	M	I - III
15	4	1435	143.57	35.89	38	Severe	Pain	F	IV - VI
16	16	2438	243.61	15.23	66	None	No Pain	M	I - III
17	33	12556	1253.93	38	328	None	No Pain	F	I - III
18	3	2779	278.08	92.69	71	None	No Pain	M	I - III
19	18	635	63.38	3.52	21	Severe	Pain	F	I - III
20	8	3222	322.18	40.27	84	Low	Pain	M	I - III
21	50	14904	1489	29.78	397	Moderate	Pain	F	I - III
22	11	5413	541.28	49.21	141	Severe	Pain	F	I - III
23	9	3711	371.34	41.26	98	Severe	Pain	M	IV - VI
24	17	5332	533.44	31.38	142	None	No Pain	F	I - III
25	15	3504	350.6	23.37	95	None	No Pain	M	IV - VI
26	4	1073	107.43	26.86	28	Severe	Pain	M	IV - VI
27	9	1478	147.93	16.44	40	None	No Pain	M	IV - VI
28	19	1467	146.48	7.71	44	Low	Pain	M	IV - VI
29	6	1933	193.27	32.21	51	None	No Pain	M	IV - VI
30	4	1697	169.73	42.43	45	Low	Pain	M	IV - VI

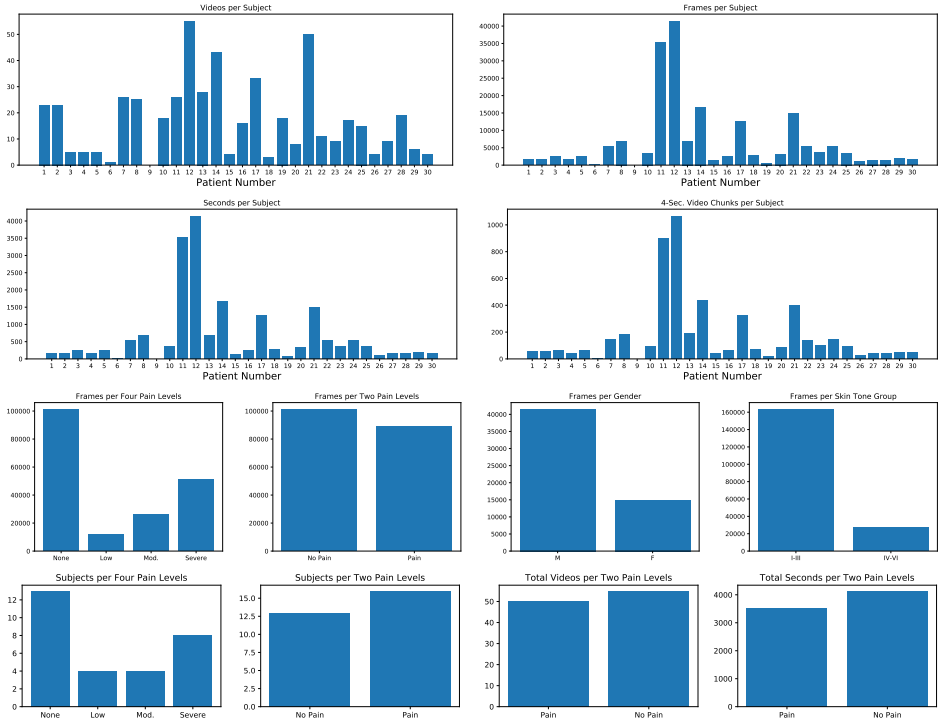


Figure 9: **Plots of Descriptive Statistics**. Plots are consistent with Table 6 shown above.

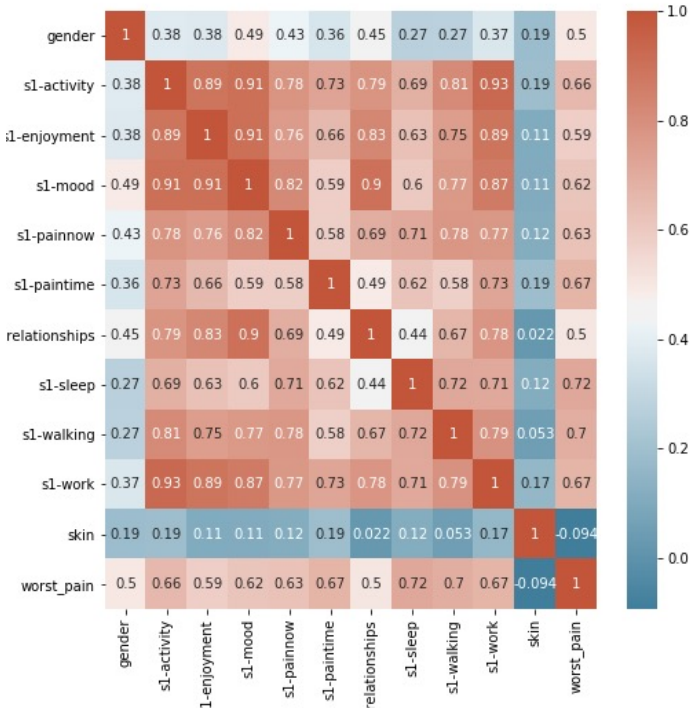


Figure 10: **Correlation Analysis of Self-Reported Pain Scores**

Table 7: **Source Devices Used to Submit Data by each Patient**

Patient	Source Device
0001	Android
0002	iOS
0003	iOS
0004	iOS
0005	Clinic Only
0006	Clinic Only
0007	Android
0008	iOS
0009	Android
0010	iOS
0011	iOS
0012	Android
0013	Android
0014	iOS
0015	iOS
0016	Android
0017	Android
0018	iOS
0019	Android
0020	iOS
0021	Android
0022	iOS
0023	iOS
0024	iOS
0025	iOS
0026	Android
0027	iOS
0028	iOS
0029	iOS

F Details of the Models

F.1 Cross Validation Details

We use 10-fold-cross-validation leading to 10 different training splits, where 9 of the splits reserve 3 patients for the test set and the tenth split uses 2 patients, since there are 29 patients.

Table 8: **10-fold-CV details - Facial Image Counts per Split.**

	Face Images					
	Train Set		Test Set		Total Data	
Split	Face Images	Face Images w/LMs	Face Images	Face Images w/LMs	Face Images	Face Images w/LMs
1	157269	153322	15742	14741	173011	168063
2	163128	158310	9883	9753	173011	168063
3	163734	158809	9277	9254	173011	168063
4	162678	157865	10333	10198	173011	168063
5	156622	151809	16389	16254	173011	168063
6	168339	163578	4672	4485	173011	168063
7	145522	141804	27489	26259	173011	168063
8	103418	100462	69593	67601	173011	168063
9	166739	161816	6272	6247	173011	168063
10	169650	164792	3361	3271	173011	168063

Table 9: **10-fold-CV details - 4-Second Chunks of Mel Spectrograms and Audio Features, Counts per Split.**

Split	Train Set	Test Set	Total
1	4486	502	4988
2	4719	269	4988
3	4742	246	4988
4	4710	278	4988
5	4561	427	4988
6	4858	130	4988
7	4264	724	4988
8	2841	2147	4988
9	4821	167	4988
10	4890	98	4988

Table 10: **10-fold-CV details - Pain Scores that include Skin, sex, and Timeframe Labels, Counts per Split.**

Split	Train Set	Test Set	Total
1	455	82	537
2	479	58	537
3	505	32	537
4	511	26	537
5	497	40	537
6	495	42	537
7	449	88	537
8	425	112	537
9	510	27	537
10	507	30	537

Table 11: **10-fold-CV details - Test Patients per Split.** We show each patient held out in the test set for each split. For each test patient, the table shows the original pain level and the binary pain level in order of the patient. For example, for Split 1 Test Patient 0002, the original pain level is “Low" and the binary pain level is “Pain".

Split	Test Patients			Respective Original Pain Level			Respective Binary Pain Level		
1	0002	0029	0021	Low	None	Moderate	Pain	No Pain	Pain
2	0028	0027	0008	Low	None	Moderate	Pain	No Pain	Pain
3	0020	0025	0005	Low	None	Moderate	Pain	No Pain	Pain
4	0015	0024	0023	Severe	None	Severe	Pain	No Pain	Pain
5	0018	0017	0026	None	None	Severe	No Pain	No Pain	Pain
6	0004	0016	0019	Moderate	None	Severe	Pain	No Pain	Pain
7	0014	0022	0007	None	Severe	Low	No Pain	Pain	No Pain
8	0013	0011	0012	None	Severe	None	No Pain	Pain	No Pain
9	0003	0010	0006	None	Severe	Severe	No Pain	Pain	Pain
10	0030	0001		None	None		No Pain	No Pain	

F.2 Model Architectures

Diagrams provide model architecture details for Experiments 4 - 7. The architecture for Experiment 1 is ResNet50, unmodified, pre-trained on ImageNet and fine-tuned. Models used for Experiments 2 and 3 are Random Forest Classifiers using all default settings in the `sklearn.ensemble.RandomForestClassifier` library with 100 estimators

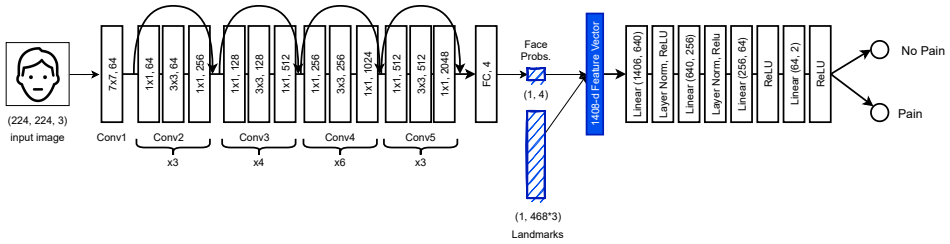


Figure 11: Experiment 4: Fusion 1, Faces + Landmarks.

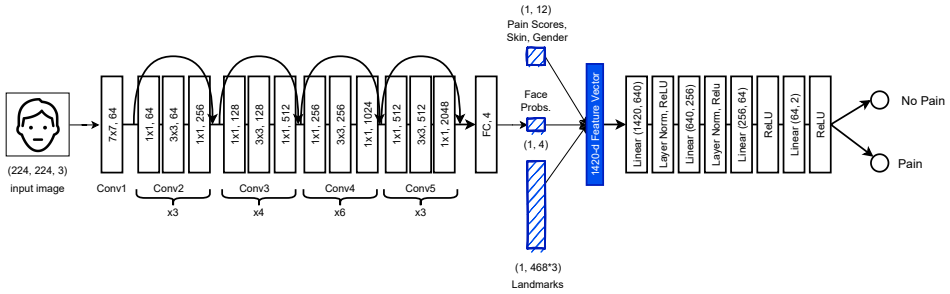


Figure 12: Experiment 5: Fusion 2, Faces + Landmarks + Pain Scores.

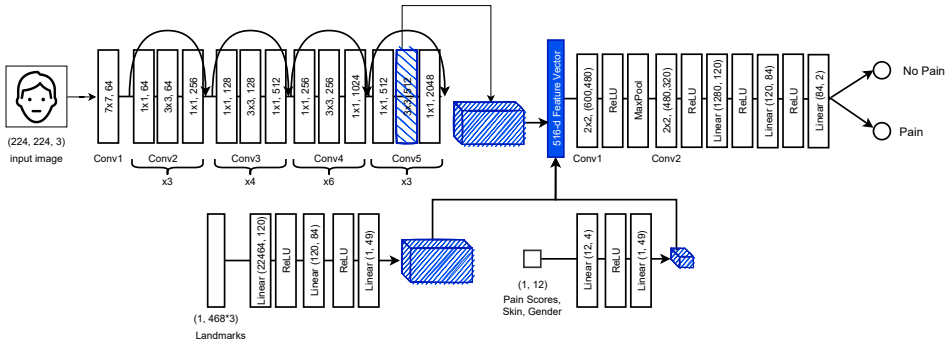


Figure 13: Experiment 6: Fusion 3, Faces + Landmarks + Pain Scores.

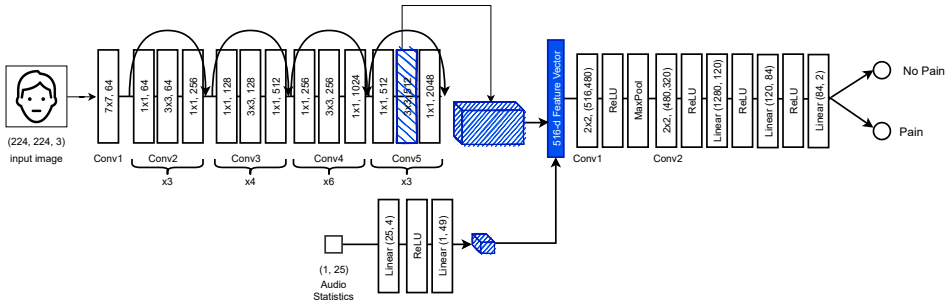


Figure 14: Experiment 7: Static Audio, Audio only.

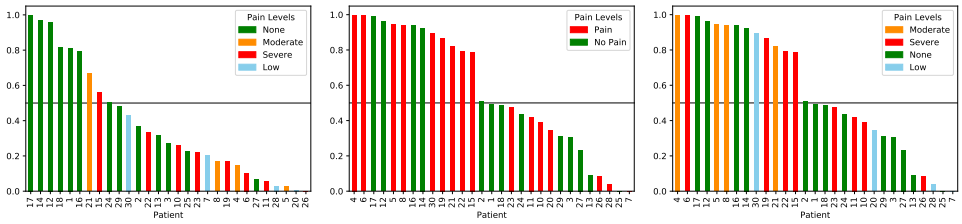
G Additional Model Results

We provide additional results from baseline experiments. In Table 12, the number and percentage of patients with accuracy scores in the ranges of 0 - 25%, 25% - 50%, 50% - 75%, and 75% is shown. In Figure 15, we show the test patient accuracy scores by binary labels as well as the original four pain labels. There is an imbalance of pain levels in the ISS dataset, where “None" frames outnumber “Low" and “Moderate" classes at a ratio of 8.50:1 and 3.88:1, respectively. When inspecting the binary classifier according to the original four pain

labels, the “Moderate" test patients show greater accuracy. In Figure 16 we see a decrease in the accuracy of the “Moderate" and “Low" test patients. In Figure 17, while the “Moderate" classes exceed 50% accuracy, the model struggles with “Low" pain test patients. However, 13 test patients reach test accuracies close to 1.0.

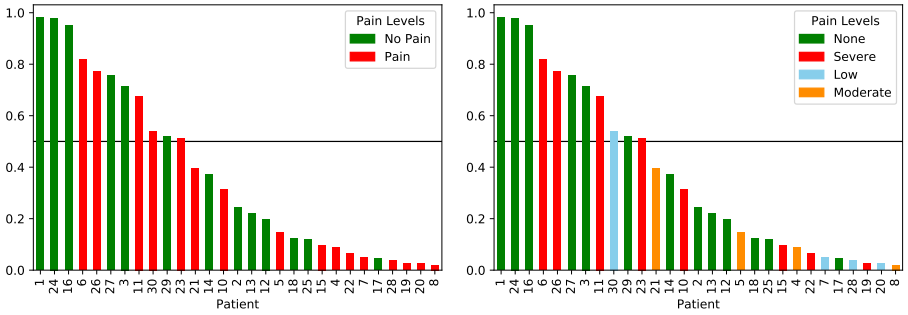
Table 12: **Model Results by Patient Scores by Accuracy Range.**

		Test Accuracy Ranges									
		[0. 0.25)		[0.25, 0.50)		[0.50, 0.75)		[0.75, 1.0)		Over 50%	
Experiments	**Model**	**No. Patients**	**Perc. Patients**	**No. Patients**	**Perc. Patients**	**No. Patients**	**Perc. Patients**	**No. Patients**	**Perc. Patients**	**No. Patients**	**Perc. Patients**
Exp. 1	ResNet50-4-static	13	44.8%	7	24.1%	3	10.3%	6	20.7%	9	0.31034483
Exp. 1	ResNet50-2-static	6	20.7%	9	31.0%	1	3.4%	13	44.8%	14	0.48275862
Exp. 2	Random Forest LM	15	51.7%	3	10.3%	5	17.2%	6	20.7%	11	0.37931034
Exp. 3	Random Forest Pain	7	24.1%	2	6.9%	4	13.8%	15	**51.7%**	19	0.65517241
Exp. 4	Fusion 1	11	37.9%	3	10.3%	4	13.8%	11	37.9%	15	0.51724138
Exp. 5	Fusion 2	8	27.6%	3	10.3%	3	10.3%	15	**51.7%**	18	0.62068966
Exp. 6	Fusion 3	7	24.1%	1	3.4%	6	20.7%	15	**51.7%**	21	**0.72413793**
Exp. 7	Static Audio	9	31.0%	7	24.1%	7	24.1%	6	20.7%	13	0.44827586



(a) ResNet50-4-static (b) ResNet50-2-static - Binary (c) ResNet50-2-static - 4 Labels

Figure 15: **Accuracy per Patient for Static Face Models.** Patients are ranked in descending order of test accuracy and color coded by pain labels. (a) shows the results of Experiment 1 (ResNet50-4-static). (b) Experiment 1 (ResNet50-2-static) binary pain classification results. (c) Experiment 1 (ResNet50-4-static) results but visualized with the original four pain labels.

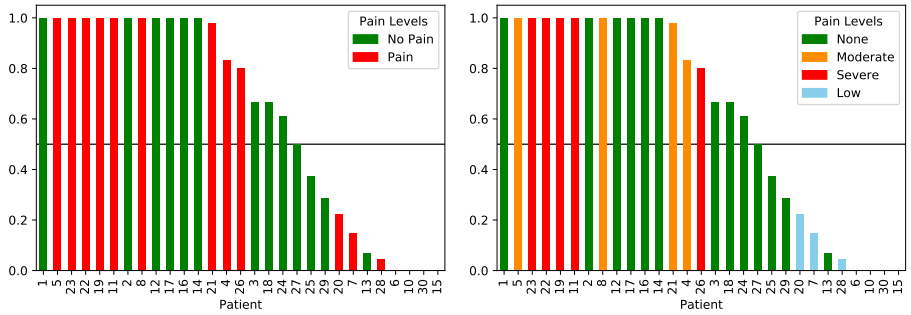


(a) Random Forest LM - Binary (b) Random Forest LM - 4 Labels

Figure 16: **Accuracy per Patient for Landmark Models.**Patients are ranked in descending order of test accuracy and color coded by pain labels. (a) shows the results of Experiment 2 (Random Forest LM) binary pain classification results. (b) shows Experiment 2 (Random Forest LM) results but visualized with the original four pain labels.

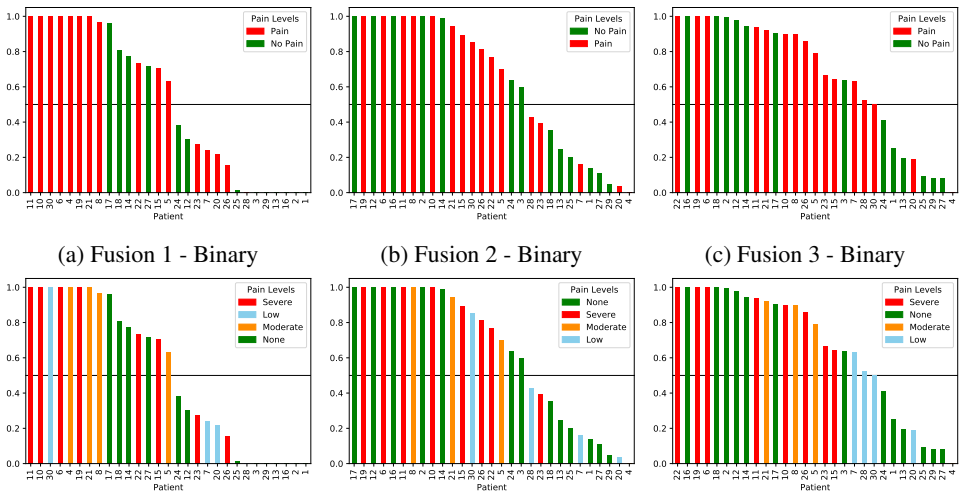
Table 13: Bottom Five Test Patients for Each Model, with Near Zero Accuracy.

Experiments	**Model**	**Bottom 5 Test Patients**
Exp. 1	Resnet50-2-static	7, 25, 28, 26, 13
Exp. 2	Random Forest LM	8, 20, 19, 28, 17
Exp. 3	Random Forest Pain	15, 30, 10, 6, 28
Exp. 4	Fusion 1	1, 2, 16, 13, 29, 25*
Exp. 5	Fusion 2	4, 20, 29, 17, 1
Exp. 6	Fusion 3	4, 27, 29, 25, 20



(a) Random Forest Pain - Binary (b) Random Forest Pain - 4 Labels

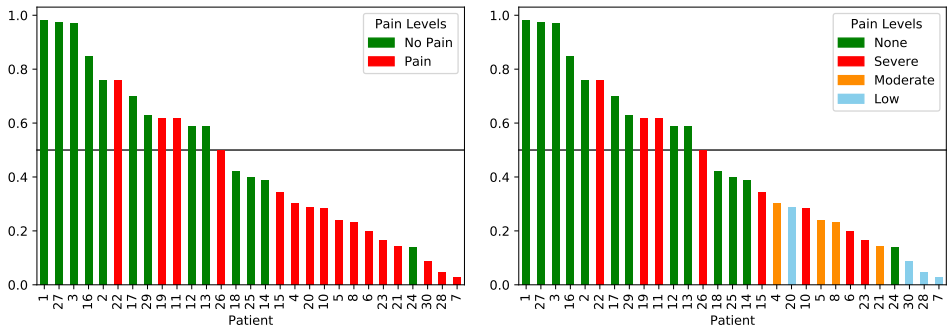
Figure 17: **Accuracy per Patient for Pain Models.** Patients are ranked in descending order of test accuracy and color coded by pain labels. (a) shows the results of Experiment 3 (Random Forest Pain) binary pain classification results. (b) shows Experiment 3 (Random Forest Pain) results but visualized with the original four pain labels.



(a) Fusion 1 - Binary (b) Fusion 2 - Binary (c) Fusion 3 - Binary

(d) Fusion 1 - 4 Labels (e) Fusion 2 - 4 Labels (f) Fusion 3 - 4 Labels

Figure 18: **Accuracy per Patient for Multimodal Face, Landmark and Pain Score Models.** Each subplot shows the results of Experiments 4, 5, and 6, with the binary predictions as well as visualized with the original four pain levels to provide greater granularity. Plots (a) and (d) provide accuracy scores for each test patient for Experiment 4 that combines facial images and landmarks. Plots (b) and (e) provide accuracy scores for Experiment 5 that combines facial images, raw landmarks, and raw pain scores, with sex, skin tone, and timeframe labels by simple concatenations. Plots (c) and (f) provide results of Experiment 6 that combines learned features from separate networks for each modality of the facial images, landmarks, and pain scores with sex, skin tone, and timeframe labels, into a single multimodal network.



(a) Audio, 2 Labels (b) Audio, 4 Labels

Figure 19: **Accuracy per Patient for Audio Models.**

H Datasheets for Datasets

H.1 Motivation

For what purpose was the dataset created? The ISS dataset was created to explore whether facial video, images, audio, text, and pain scores, could be used to predict chronic cancer self-reported pain across a diverse patient group. [Paper section reference: Introduction]

Who created the dataset? The National Institutes of Health (NIH) National Cancer Institute (NCI). [Paper section reference: Introduction]

Who funded the creation of the dataset? The National Institutes of Health (NIH) National Cancer Institute (NCI). [Paper section reference: Introduction]

Any other comments? [N/A]

H.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. The ISS dataset consists of 29 cancer patients currently undergoing active treatment at the NIH, scoring their self-reported pain levels and narrating their feelings about pain. [Paper section reference: ISS Dataset/Sample and Study Design]

How many instances are there in total (of each type, if appropriate)? 29 patients are currently enrolled in the study. [Paper section reference: ISS Dataset/Sample and Study Design]

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? It is the current number of enrolled patients to date. The larger set will encompass 112 patients once the study has concluded. [Paper section reference: ISS Dataset/Sample and Study Design]

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? Each patient submits a video through their smartphone or home computer. The video is extracted into video frames, facial images, facial landmarks, audio files, spectrograms, and audio features. [Paper section reference: ISS Dataset/Data Description]

Is there a label or target associated with each instance? Yes, the label is "worst pain" which represents the patient's pain level reported in the past 30 days upon enrollment in the study. This label does not change through the patient's time in the study. [Paper section reference: ISS Dataset/Sample and Study Design; Baselines]

Is any information missing from individual instances? Yes, Patient 0015 lacks self-reported pain scores, and Patient 0009 video data was unusable. [Paper section reference: ISS Dataset/Sample and Study Design]

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? N/A.

Are there recommended data splits (e.g., training, development/validation, testing)? Yes. We use 10-fold-cross-validation that withholds three patients from nine of the splits, and two patients from the tenth split. [Paper section reference: Appendix/Details of the Models/Cross Validation Details]

Are there any errors, sources of noise, or redundancies in the dataset? Patients are inconsistent in their video recording, leading to noisy backgrounds such as low illumination, muffled or quiet speaking, letters and signage in the background, shiny bright sunlight and glare, and occasional mask wearing due to Covid-19. [Paper section reference: Discussion and Future Work]

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g.,websites, tweets, other datasets)? The ISS dataset will be self-contained and hosted under the supervision of the NIH. [Paper section reference: ISS Dataset/Data Storage and Access]

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)? Yes. The dataset consists of PII in the form of people's video, faces, and medical discussions. [Paper section reference: ISS Dataset/Data Storage and Access]

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? Unknown.

Does the dataset relate to people? Yes, it consists of 29 cancer patients undergoing active treatment. [Paper section reference: ISS Dataset/Sample and Study Design]

Does the dataset identify any subpopulations (e.g., by age, sex)? The dataset is grouped into cohorts assigned by age and sex. [Paper section reference: ISS Dataset/Sample and Study Design]

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? No. An anonymous identifier has been used to label each patient. The dataset has been deidentified of patient name, date of birth, address, or medical record number. [Paper section reference: ISS Dataset/Data Storage and Access; Appendix/Author Statement]

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial 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)? No.

Any other comments? N/A

H.3 Collection Process

How was the data associated with each instance acquired? Each patient submits a video through a mobile or web-based application. [Paper section reference: ISS Dataset/Patient Protocol]

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? A mobile/web-based application is the main interface used to collect the self-reported pain scores and video recording. Each patient uses their own smartphone or computer in their personal home setting to submit the video. [Paper section reference: ISS Dataset/Patient Protocol]

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

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)? Compensation for patients enrolled in the study is explained in the Consent Form as follows ([Paper section reference: Appendix/Consent Form]):

- $15 per week in which you successfully submit three (3) or more completed questionnaires, including self-video recordings. A maximum of one (1) questionnaire may be completed per day.
- $15 per in-clinic questionnaire and video recording session successfully completed. A minimum of one in-clinic recording must be completed within three months.
- No matter how many questionnaires or in-clinic sessions you complete, the total maximum allowed compensation for this study is $225.

Over what timeframe was the data collected? Patients reported in this dataset were enrolled in the study between December 2020 and July 2021. [Paper section reference: ISS Dataset/Sample and Study Design; Related Works Table 1]

Were any ethical review processes conducted (e.g., by an institutional review board)? Yes, NCI protocol 20C0130 entitled, "A Feasibility Study Investigating the Use of Machine Learning to Analyze Facial Imaging, Voice and Spoken Language for the Capture and Classification of Cancer/Tumor Pain." `https://clinicaltrials.gov/ct2/show/NCT04442425` [Paper section reference: Introduction]

Does the dataset relate to people? Yes.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)? From the individuals directly. [Paper section reference: ISS Dataset/Patient Protocol]

Were the individuals in question notified about the data collection? Yes. ([Paper section reference: Appendix/Consent Form])

Did the individuals in question consent to the collection and use of their data? Yes. [Paper section reference: Appendix/Consent Form]

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? Yes. Each patient signed an NIH Assent to Participate in a Clinical Research Study, reviewed alongside their physician which provides information about the study such as its purpose, requirements, benefits, and monetary incentives. The form indicates that the patient may change their mind at any point in time and drop out of the study. [Paper section reference: ISS Dataset/Sample and Study Design; Appendix/Consent Form]

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? Yes, during the IRB approval process.

Any other comments? N/A

H.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)? Yes. Video data was extracted for video frames at 10 frames per second, faces cropped using a facial detection algorithm, and landmarks (AAMs) detected. Video data was extracted form audio, and the resulting audio file was used to generate a Mel spectrogram and audio features. Minimal transformations were applied to the extracted data. [Paper section reference: ISS Dataset/Data Description; Appendix/Preprocessing]

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? Yes.

Is the software used to preprocess/clean/label the instances available? Yes, we use open source Python libraries to include OpenCV for image processing, ffmpeg for video extraction, PyDub for audio file chunking, Librosa for audio feature extraction and Mel spectrogram generation, Scikit-Learn for machine learning models, and PyTorch for neural networks. [Paper section reference: ISS Dataset/Data Description; Appendix/Preprocessing]

Any other comments? N/A

H.5 Uses

Has the dataset been used for any tasks already? No.

Is there a repository that links to any or all papers or systems that use the dataset? No.

What (other) tasks could the dataset be used for? Our baseline experiments only addressed one task of static frame-by-frame pain classification. However, there are a variety of tasks specific to improving self-reported chronic pain prediction, that can be explored with the ISS dataset. They are:

- Temporal-based, video classification model, integrating image, audio, and text.
- Disentangling the model's dependency on patients and their identity.
- Predicting multiple features of pain such as the average self-reported affective and activity scores.
- Development and automation of chronic cancer pain-specific facial action units.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? No.

Are there tasks for which the dataset should not be used? General emotion detection, facial recognition, surveillance applications.

Any other comments? N/A

H.6 Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description. No. NIH will maintain guardianship of the data.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)? To be determined.

When will the dataset be distributed? To be determined, under the guidance of the NIH.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions. To be determined, under the guidance of the NIH. [Paper section reference: Appendix/Author Statement]

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions. No.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation. No.

Any other comments? Prior to accessing the data, it is likely that users will be required to provide proof of successful completion of some or all of the following types of training: Protection of Human Subjects in Research, Global Privacy and Data Protection, a HIPAA and Health Privacy, General Information Security, Social and Behavioral Research, and/or Conflicts of Interest.

H.7 Maintenance

Who will be supporting/hosting/maintaining the dataset? The NIH. [Paper section reference: ISS Dataset/Data Storage and Access]

How can the owner/curator/manager of the dataset be contacted (e.g., email address)? `gulleyj@mail.nih.gov`

Is there an erratum? If so, please provide a link or other access point. N/A.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)? To be determined, under the guidance of the NIH.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced. Patients were informed through the Consent Form that their data may be held indefinitely. [Paper section reference: Appendix/Consent Form]

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to users. To be determined, under the guidance of the NIH.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description. To be determined, under the purview of the NIH.

Any other comments? N/A.

I Risk Categorization

Our study was classified as "Minimal Risk under 45 CFR 46 / 21 CFR 56" after review and approval by the National Institutes of Health's Federalwide Assurance (FWA00005897), in accordance with Federal regulations 45 CFR 46 and 21 CFR 56, including Good Clinical Practice. The study also was determined to fall under three "Expedited Review Categories" due to three applicable characteristics including (1b) clinical studies of medical devices for which an investigational device exemption application (21 CFR Part 812) is not required, (6) collection of data from voice, video, digital, or image recording made for research purposes, and (7) research on individual or group characteristics or behavior. The data collection which involved the use of a mobile or web app to input answers to questions and record videos do not pose risk to patients. The tasks also do not involve deception.

PRINCIPAL INVESTIGATOR: James Gulley, MD, PhD

STUDY TITLE: A Feasibility Study Investigating the Use of Machine Learning to Analyze Facial Imaging, Voice and Spoken Language for the Capture and Classification of Cancer/Tumor Pain

STUDY SITE: NIH Clinical Center (CC)

Cohort: *Affected Participants*

Consent Version: *05.21.2021*

WHO DO YOU CONTACT ABOUT THIS STUDY?

Principal Investigator: James Gulley, MD, PhD
Phone: 301-480-7164
Email: james.gulley@nih.gov

KEY INFORMATION ABOUT THIS RESEARCH

This consent form describes a research study and is designed to help you decide if you would like to be a part of the research study.

You are being asked to take part in a research study at the National Institutes of Health (NIH). This section provides the information we believe is most helpful and important to you in making your decision about participating in this study. Additional information that may help you decide can be found in other sections of the document. Taking part in research at the NIH is your choice.

You are being asked to take part in this study because you have cancer or a tumor for which you are actively undergoing treatment.

The purpose of this research study is to find out whether we can use facial recognition technology to classify pain in a diverse set of patients with cancer or tumors. Another purpose of this study is to find out whether we can use voice recognition technology to transcribe patient video responses to assess pain.

The facial and recognition technology used on this study is not approved by the FDA, but the FDA allows us to use the technology in clinical trials such as this one.

You will not receive any treatment for your cancer or tumor on this study.

If you decide to join this study, here are some of the most important things that you should know that will happen:

- You will be asked to come to the NIH Clinical Center between one and four times while you are participating in this study. You will also be asked to check-in at-home approximately 3 times per week during the study. The assessments you will take part in at the Clinical Center (in-clinic) or remotely (at-home) will take about 3 minutes to complete.

- During these check-ins you will be asked to:
  - complete a questionnaire in an application using a mobile phone or a computer with a camera/microphone,
  - record a video of yourself reading a 15 second passage of text, and
  - record a video of yourself responding to an additional question.

You will be on the study for about 3 months.

You will be compensated for your participation as long as you complete at least six or more total submissions (at-home or in-clinic). You will receive $15 per week in which you submit three or more questionnaires, including video recordings, completed at-home. You will receive an additional $15 for each completed series of video recordings and questionnaires completed in-clinic if you complete at least two questionnaires. You will be paid a maximum of $225 for the duration of this study. You will not be paid until you have completed the study. No compensation will be provided for the week if you complete less than three questionnaires in home or less than one in-clinic session.

You will use your own internet connection to access the software used on this study, so depending on your service plan, you may be charged by your provider.

The databases used for this study are secure, but it is possible that in spite of our efforts, someone might gain access to the data we collect. We do not expect that you will benefit from taking part in this study. However, the results from our research may help others in the future.

You are free to stop participating in the trial at any time.

The remaining document will now describe the research study in more detail. This information should be considered before you make your choice. Members of the study team will talk with you about the information in this document. Some people have personal, religious, or ethical beliefs that may limit the kinds of medical or research interventions in which they would want to participate. Take the time you need to ask any questions and discuss this study with NIH staff, and with your family, friends, and personal health care providers.

If the individual being enrolled is a minor then the term "you" refers to "you and/or your child" throughout the remainder of this document.

IT IS YOUR CHOICE TO TAKE PART IN THE STUDY

You may choose not to take part in this study for any reason. If you join this study, you may change your mind and stop participating in the study at any time and for any reason. In either case, you will not lose any benefits to which you are otherwise entitled. However, to be seen at the NIH, you must be taking part in a study or are being considered for a study. If you do choose to leave the study, please inform your study team to ensure a safe withdrawal from the research.

WHY IS THIS STUDY BEING DONE?

This is a research study. The purpose of this research study is to find out whether we can use facial recognition technology to classify pain in a diverse set of patients with cancer or tumors. Another purpose of this study is to find out whether we can use voice recognition technology to transcribe patient video responses to assess pain.

We are asking you to join this research study because you have been diagnosed with cancer or a tumor and are currently receiving active treatment for your disease.

The facial and recognition technology used on this study is not approved by the FDA, but the FDA allows us to use the technology in clinical trials such as this one.

You will not receive any treatment for your cancer or tumor on this study.

WHAT WILL HAPPEN DURING THE STUDY?

If you decide to take part in this study, you will be asked to tell us about pain related to your cancer or tumor in multiple in-clinic and at-home sessions. If there are unforeseen technical or logistical reasons that would prevent you from completing your first in-clinic check-in, you may initiate participation through remote sessions, and complete your first in-clinic check-in when practical.

During this study you will use a web based application on a mobile phone or a computer with a camera/microphone to complete a patient-reported outcomes questionnaire and record yourself completing two additional questions.

During your check-ins, both in-clinic and at-home, you will be asked to:

- complete a questionnaire in an application using a mobile phone or a computer with a camera/microphone,
- record a video of yourself reading an approximately 15 second passage of text, and
- record a video of yourself responding to an additional question(s).

During the in-clinic check-ins a study team member will use professional lighting and videography equipment, including thermal cameras, for the recorded portions of the assessments. Your first evaluation will occur at the NIH Clinical Center.

During remote check-ins, you will be asked to complete the questionnaire and recordings alone, in a quiet and brightly lit room, preferably with a white wall or background. You will use a camera on a mobile device or a computer for your recordings.

NOTE: Data that you will submit for this study, including recordings and information entered into the NCI Pain Insights application, will not be actively monitored. Participation in this study and submitting your recordings and questionnaires does not replace your direct communication with your healthcare team regarding your health issues.

HOW LONG WILL THE STUDY TAKE?

If you agree to take part in this study, your involvement is expected to last for 3 months.

You will be asked to check-in remotely approximately 3 times per week but no more than once daily.

You will be asked to check-in at the NIH Clinical Center at least 1 and up to 4 times during this study.

Your assessments (remote and in-clinic check-ins) are expected to last approximately 3 minutes, but not longer than 30 minutes.

You may continue to complete remote and in-clinic check-ins for this study even if you are taken off-treatment or off-study on the primary NIH treatment protocol.

HOW MANY PEOPLE WILL PARTICIPATE IN THIS STUDY?

We plan to have approximately 120 people participate in this study at the NIH.

WHAT ARE THE RISKS AND DISCOMFORTS OF BEING IN THE STUDY?

It is possible that in spite of our efforts, the information we collect about you may be seen by persons not authorized to see it. However, the safeguards in place should mean that there is a very small chance that this will occur. There are no other known possible risks associated with participation in this study. You will respond to questions using keyboard and audio-visual recording devices.

WHAT ARE THE BENEFITS OF BEING IN THE STUDY?

You will not benefit from being in this study.

Are there any potential benefits to others that might result from the study?

In the future, other people might benefit from this study because we may be able to determine pain from cancer or tumors using facial recognition technology.

WHAT OTHER OPTIONS ARE THERE FOR YOU?

The alternative to participation in the study is to not participate. While enrolled on this study, you will remain under the care of your physician and under another protocol of the National Institutes of Health for your cancer or tumor treatment.

DISCUSSION OF FINDINGS

New information about the study

If we find out any new information that may affect your choice to participate in this study, we will get in touch with you to explain what we have learned. This may be information we have learned while doing this study here at the NIH or information we have learned from other scientists doing similar research in other places.

Return of research results

No research results will be returned to you.

EARLY WITHDRAWAL FROM THE STUDY

Your doctor may decide to stop your participation in this study for the following reasons:

- if he/she believes that it is in your best interest;
- if you lose your ability to give consent
- if the "NCI Pain Insights" application becomes unavailable;
- if you do not follow the study rules;
- if the study is stopped for any reason.

In this case, you will be informed of the reason therapy is being stopped.

You can stop taking part in the study at any time.

If you decide at any time to withdraw your consent to participate in the trial, we will not collect any additional medical information about you. However, according to FDA guidelines,

information collected on you up to that point may still be provided to Booz Allen Hamilton or designated representatives.

STORAGE, SHARING AND FUTURE RESEARCH USING YOUR DATA

Will Your Data Be Saved for Use in Other Research Studies?

As part of this study, we are obtaining data from you. We will remove all the identifiers, such as your name, date of birth, address, or medical record number and label your data with a code so that you cannot easily be identified. However, the code will be linked through a key to information that can identify you. We plan to store and use these data for studies other than the ones described in this consent form that are going on right now, as well as studies that may be conducted in the future. These studies may provide additional information that will be helpful in understanding cancer, or other diseases or conditions. This could include studies to develop other research tests, treatments, drugs, or devices, that may lead to the development of a commercial product by the NIH and/or its research or commercial partners. There are no plans to provide financial compensation to you if this happens. Also, it is unlikely that we will learn anything from these studies that may directly benefit you.

I give permission for my coded data to be stored and used for future research as described above.

_____ Yes _____ No

Initials Initials

Will Your Data Be Shared for Use in Other Research Studies?

We may share your coded data with other researchers. If we do, while we will maintain the code key, we will not share it, so the other researchers will not be able to identify you. They may be doing research in areas that are similar to this study or in other unrelated areas. These researchers may be at NIH, other research centers and institutions, or commercial entities.

I give permission for my coded data to be shared with other researchers and used by these researchers for future research as described above.

_____ Yes _____ No

Initials Initials

If you change your mind and do not want us to store and use your data for future research, you should contact the research team member identified at the top of this document. We will do our best to comply with your request but cannot guarantee that we will always be able to destroy your data. For example, if some research with your data has already been completed, the information from that research may still be used. Also, for example, if the data have been shared already with other researchers, it might not be possible to withdraw them.

In addition to the planned use and sharing described above, we might remove all identifiers and codes from your data and use or share them with other researchers for future research at the NIH or other places. When we or the other researchers access your anonymized data, there will be no

way to link the data back to you. We will not contact you to ask your permission or otherwise inform you before we do this. We might do this even if you answered "no" to the above questions. If we do this, we would not be able to remove your data to prevent their use in future research studies, even if you asked, because we will not be able to tell which are your data.

NIH policies require that your clinical and other study data be placed in an internal NIH database that is accessible to other NIH researchers for future research. Usually, these researchers will not have access to any of your identifiers, such as your name, date of birth, address, or medical record number; and your data will be labeled with only a code. We cannot offer you a choice of whether your data to be placed in this database or not. If you do not wish to have your data placed in this database, you should not enroll in this study.

How Long Will Your Data be Stored by the NIH?

Your data may be stored by the NIH possibly indefinitely.

Risks of Storage and Sharing of Data

When we store your data, we take precautions to protect your information from others that should not have access to it. When we share your data, we will do everything we can to protect your identity, for example, when appropriate, we remove information that can identify you. Even with the safeguards we put in place, we cannot guarantee that your identity will never become known or someone may gain unauthorized access to your information. New methods may be created in the future that could make it possible to re-identify your data.

COMPENSATION, REIMBURSEMENT, AND PAYMENT

Will you receive compensation for participation in the study?

Some NIH Clinical Center studies offer compensation for participation in research. The amount of compensation, if any, is guided by NIH policies and guidelines.

You will be compensated for you participation as follows:

- You will receive $15 per week in which you successfully submit three (3) or more completed questionnaires, including self-video recordings. A maximum of one (1) questionnaire may be completed per day.
- You will receive $15 per in-clinic questionnaire and video recording session successfully completed. A minimum of one in-clinic recording must be completed within three months. You will be compensated for a maximum of four in-clinic sessions during the study.
- No matter how many questionnaires or in-clinic sessions you complete, the total maximum allowed compensation for this study is $225.

You will be paid at the end of the three months participation.

- You must complete a minimum of three submissions from home in one week or at least one in-clinic session to be eligible for compensation that week. If you do not complete these minimum submissions, either at-home or in-clinic, you will not receive any compensation for your participation that week.

If you are unable to finish the study, and have completed at least six submissions, you will receive compensation as described above for the parts you completed. If you are unable to finish the study and you have completed less than six submissions, you will not receive any compensation.

With few exceptions, study compensation is considered taxable income that is reportable to the Internal Revenue Service (IRS). A "Form 1099-Other Income" will be sent to you if your total payments for research participation are $600 or more in a calendar year. If you have unpaid debt to the federal government, please be aware that some or all of your compensation may be automatically reduced to repay that debt on your behalf.

Will you receive reimbursement or direct payment by NIH as part of your participation?

Some NIH Clinical Center studies offer reimbursement or payment for travel, lodging or meals while participating in the research. The amount, if any, is guided by NIH policies and guidelines.

On this study, the NCI will cover the cost for some of your expenses. Some of these costs may be paid directly by the NIH and some may be reimbursed after you have paid. The amount and form of these payments are determined by the NCI Travel and Lodging Reimbursement Policy. You will be given a summary of the policy which provides more information.

Will taking part in this research study cost you anything?

NIH does not bill health insurance companies or participants for any research or related clinical care that you receive at the NIH Clinical Center.

- Depending on your plan, you may have to pay extra to your internet or data plan provider for access to the study software

CONFLICT OF INTEREST (COI)

The National Institutes of Health (NIH) reviews NIH staff researchers at least yearly for conflicts of interest. This process is detailed in a COI Guide. You may ask your research team for a copy of the COI Guide or for more information. Members of the research team who do not work for NIH are expected to follow these guidelines or the guidelines of their home institution, but they do not need to report their personal finances to the NIH.

Booz Allen Hamilton is providing the "NCI Pain Insights" application for this study to NIH without charge. No NIH investigator involved in this study receives payments or other benefits from any company whose drug, product or device is being tested. However, there are some research partners not associated with the NIH working on this study who may receive payments or benefits, limited by the rules of their workplace.

CLINICAL TRIAL REGISTRATION AND RESULTS REPORTING

A description of this clinical trial will be available on http://www.ClinicalTrials.gov, as required by U.S. Law. This Web site will not include information that can identify you. At most, the Web site will include a summary of the results. You can search this Web site at any time.

CONFIDENTIALITY PROTECTIONS PROVIDED IN THIS STUDY

Some of your health information, and/or information from this study will be kept in a central database for research. Your name or contact information will not be put in the database. Your test results will be identified by a unique code and the list that links the code to your name will

be kept separate from your health information. Your information may be given out if required by law. For example, certain states require doctors to report to health boards if they find a disease like tuberculosis. However, the researchers will do their best to make sure that any information that is released will not identify you.

Will your medical information be kept private?

We will do our best to make sure that the personal information in your medical record will be kept private. However, we cannot guarantee total privacy. Organizations that may look at and/or copy your medical records for research, quality assurance, and data analysis include:

- The NIH and other government agencies, like the Food and Drug Administration (FDA), which are involved in keeping research safe for people.
- National Institutes of Health Intramural Institutional Review Board
- Qualified representatives from Booz Allen Hamilton, the company who will develop the "NCI Pain Insights" application for use in this study.

The researchers conducting this study and the NIH follow applicable laws and policies to keep your identifying information private to the extent possible. However, there is always a chance that, despite our best efforts, your identity and/or information about your participation in this research may be inadvertently released or improperly accessed by unauthorized persons.

In most cases, the NIH will not release any identifiable information collected about you without your written permission. However, your information may be shared as described in the section of this document on sharing of data, and as further outlined in the following sections.

Further, the information collected for this study is protected by NIH under a Certificate of Confidentiality and the Privacy Act.

Certificate of Confidentiality

To help us protect your privacy, the NIH Intramural Program has received a Certificate of Confidentiality (Certificate). With this certificate, researchers may not release or use data or information about you except in certain circumstances.

NIH researchers must not share information that may identify you in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings, for example, if requested by a court.

The Certificate does not protect your information when it:

1. is disclosed to people connected with the research, for example, information may be used for auditing or program evaluation internally by the NIH; or
2. is required to be disclosed by Federal, State, or local laws, for example, when information must be disclosed to meet the legal requirements of the federal Food and Drug Administration (FDA);
3. is for other research;
4. is disclosed with your consent.

The Certificate does not prevent you from voluntarily releasing information about yourself or your involvement in this research.

The Certificate will not be used to prevent disclosure to state or local authorities of harm to self or others including, for example, child abuse and neglect, and by signing below you consent to those disclosures. Other permissions for release may be made by signing NIH forms, such as the Notice and Acknowledgement of Information Practices consent.

Privacy Act

The Federal Privacy Act generally protects the confidentiality of your NIH medical information that we collect under the authority of the Public Health Service Act. In some cases, the Privacy Act protections differ from the Certificate of Confidentiality. For example, sometimes the Privacy Act allows release of information from your record without your permission, for example, if it is requested by Congress. Information may also be released for certain research purposes with due consideration and protection, to those engaged by the agency for research purposes, to certain federal and state agencies, for HIV partner notification, for infectious disease or abuse or neglect reporting, to tumor registries, for quality assessment and medical audits, or when the NIH is involved in a lawsuit. However, NIH will only release information from your medical record if it is permitted by both the Certificate of Confidentiality and the Privacy Act.

POLICY REGARDING RESEARCH-RELATED INJURIES

The NIH Clinical Center will provide short-term medical care for any injury resulting from your participation in research here. In general, no long-term medical care or financial compensation for research-related injuries will be provided by the NIH, the NIH Clinical Center, or the Federal Government. However, you have the right to pursue legal remedy if you believe that your injury justifies such action.

PROBLEMS OR QUESTIONS

If you have any problems or questions about this study, or about your rights as a research participant, or about any research-related injury, contact the Principal Investigator James Gulley, MD, PhD, james.gulley@nih.gov, 301-480-8870. You may also call the NIH Clinical Center Patient Representative at 301-496-2626, or the NIH Office of IRB Operations at 301-402-3713, if you have a research-related complaint or concern.

CONSENT DOCUMENT

Please keep a copy of this document in case you want to read it again.

Adult Research Participant: I have read the explanation about this study and have been given the opportunity to discuss it and to ask questions. I consent to participate in this study.

Signature of Research Participant	Print Name of Research Participant	Date

Parent/Guardian of a Minor Participant: I have read the explanation about this study and have been given the opportunity to discuss it and to ask questions. I give permission for my child to take part in this study.

Signature of Parent/Guardian	Print Name of Parent/Guardian	Date
Signature of Parent/Guardian	Print Name of Parent/Guardian	Date

Investigator:

Signature of Investigator	Print Name of Investigator	Date