Predicting Emotion and Attention from Smartphone Behavioral Data

Dinithi Silva-Sassaman   Cat Mai   Jing Qian   Jeff Huang
dinithi_silva_-_sassaman@alumni.brown.edu   cmai@clarku.edu   jing_qian@brown.edu   jeff_huang@brown.edu
Presenter: Cat Mai '22, Clark Fall Fest 2021 | Sponsor: Professors Jeff Huang (Brown University) and Ken Basye

Introduction

The ability to detect user engagement and emotion can be useful for designing the user experience for smartphone apps. We present a usability testing toolkit to remotely record real-time **phone motion** and **back-of-phone pressure**. Two main research questions:

- **RQ1:** Who can predict user emotion and attention state better in a non-intrusive setup: user study experts or a trained ML model?
- **RQ2:** How much emotion and attention state prediction accuracy is gained with back-of-device pressure data and orientation versus only orientation data?

Previous works: Predicting attention levels using motion replay [1]; accelerometer, proximity detector, touch events [2]; facecam and eye-tracking [3]. Predicting mood using email, location, web browsing [4]; typing behaviors [5].

Motivation for back-of-phone pressure pad:

In Fig 1, in all cases, back of devices have more interaction surface to explore. Also, the two data modalities: phone motion and back-of-phone pressure are non-intrusive to users compared to video or audio recording and take less resources to process in real time.

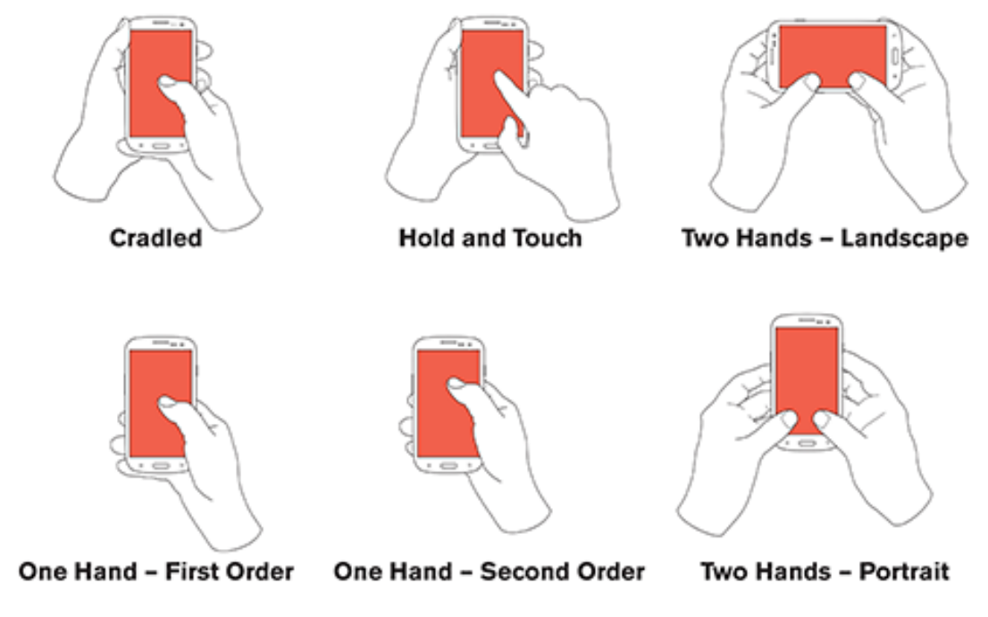


Fig 1: Common phone posture (Hoober)

Method

Ten users were recruited through university mailing list and were instructed to perform 3 tasks:

1. eBay shopping
2. Maze game
3. Dual-stick shooter game,

while their phones recorded 6dof motion and back-of-phone pressure.

Ground truth: the users were asked to self-label their data afterwards using video and audio recordings for **emotion:** *excited, relaxed, bored, frustrated* and **levels of attention:** *low, mid, high.*

We then trained a convolutional neural network (CNN) model and recruited user design experts (*analysts*). Each analyst was required to label the data from every session from every actor using Motion-only, Pressure-only, and Both data types.

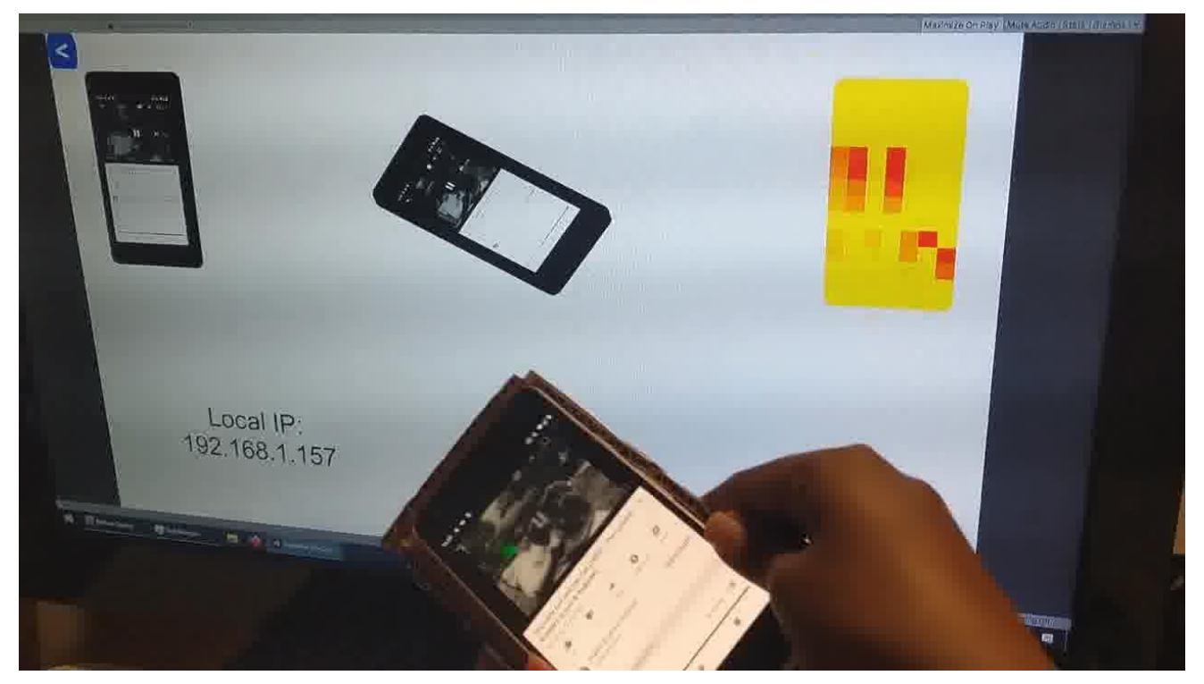


Fig 2: Motion replay is displayed through 3D model that rotates as the real device's orientation changes, as well as screen capture. The heat map varies from yellow to orange to red as the pressure increases.

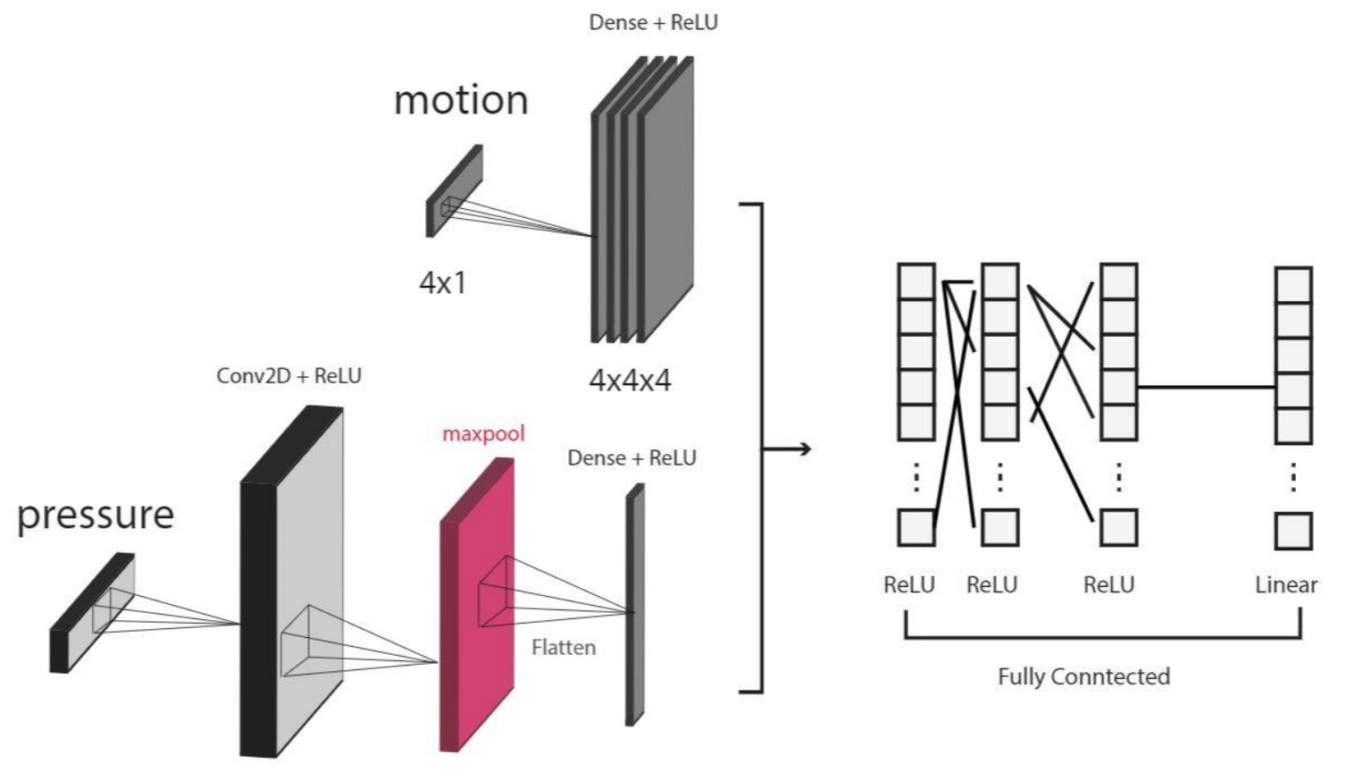


Fig 3: Layout of the CNN. Motion inputs are sent through four dense layers. Pressure inputs are processed as pictured. The two outputs are combined and sent through four extra dense layers. The last layer gives the prediction.

Results

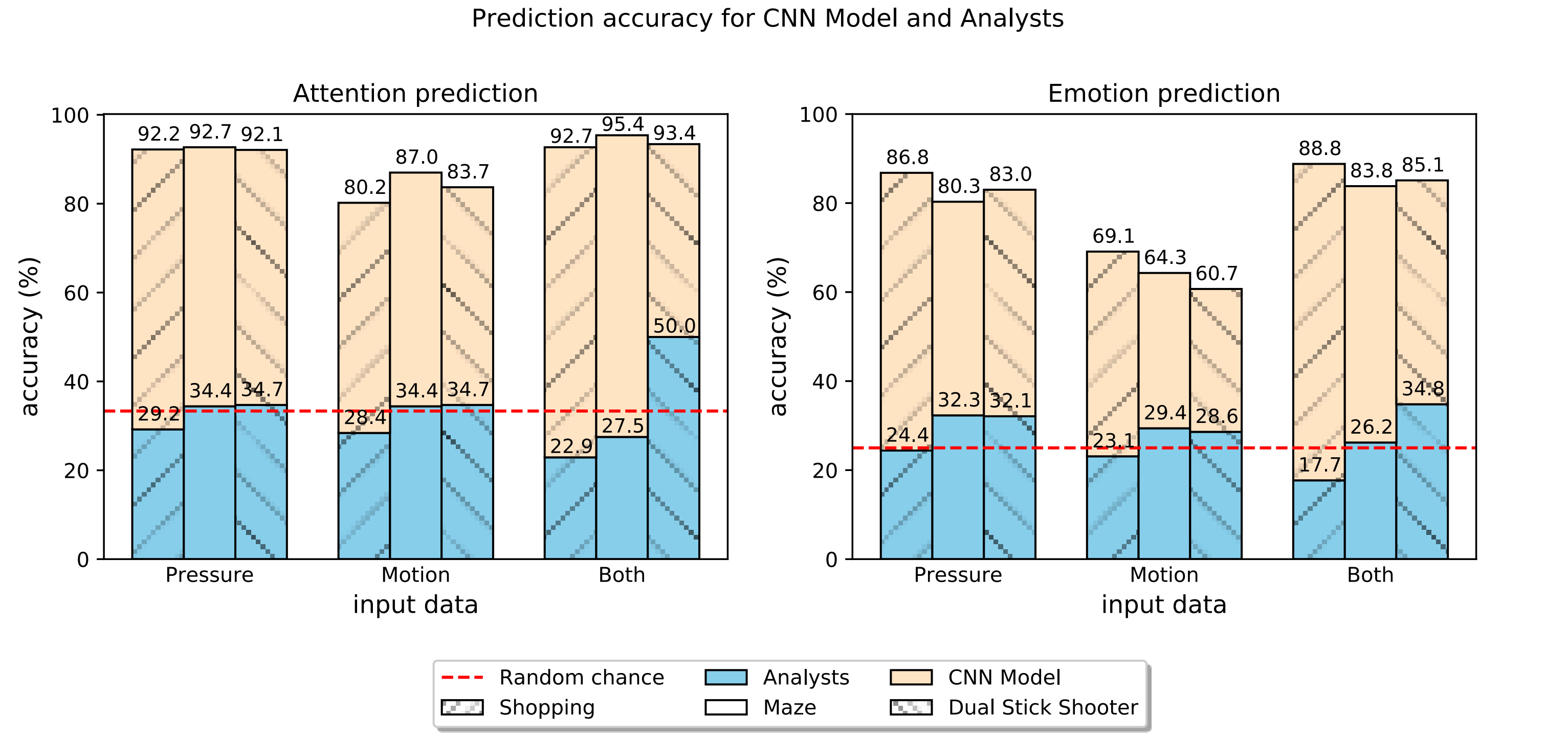
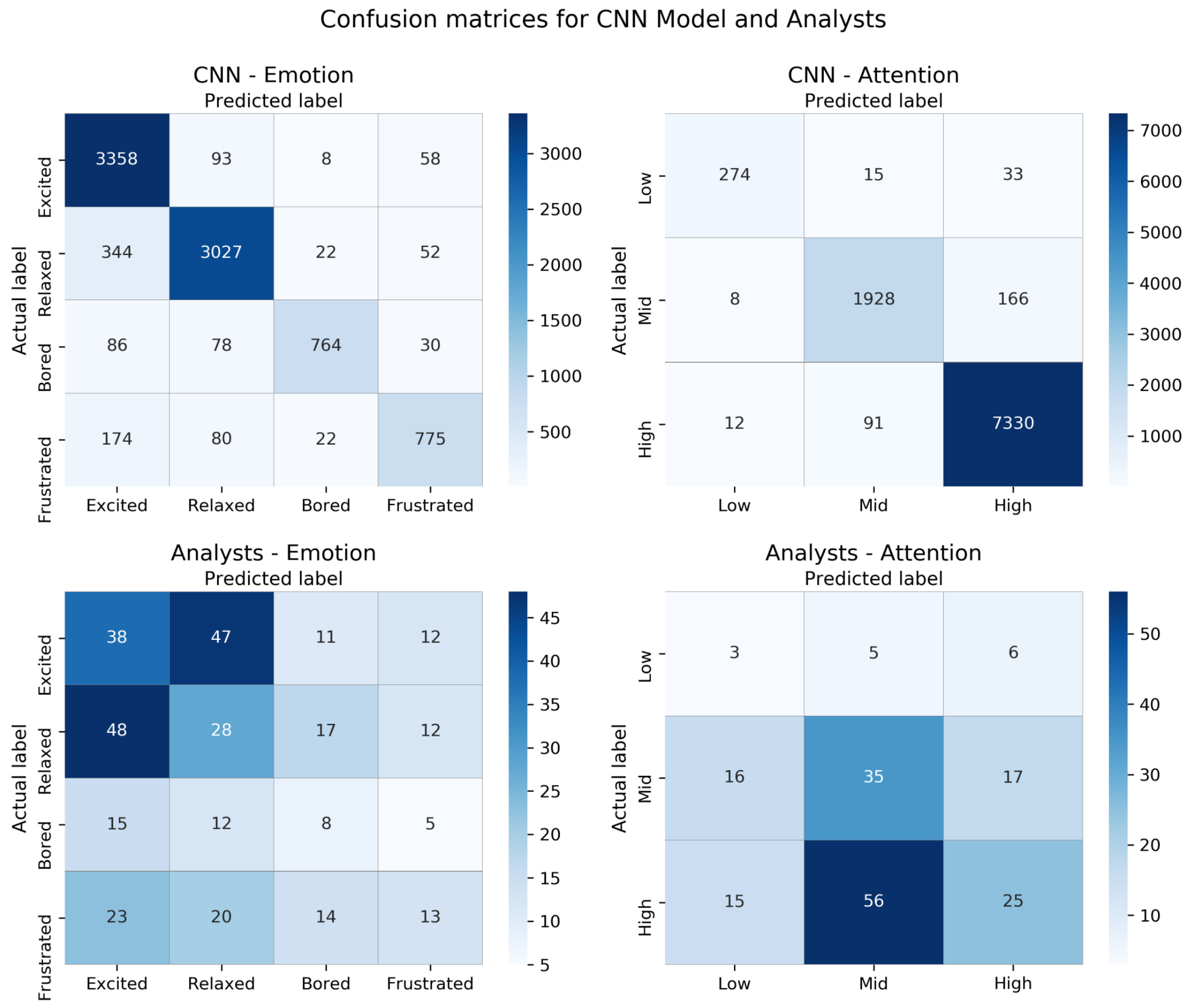


Fig 4: For each input data (Pressure, Motion, and Both), each bar represents the task being predicted. The CNN consistently outperforms analysts in both emotion and attention prediction across all tasks, regardless of input data. The red dashed line indicate a random chance to guess, which is 33%for the attention and 25% for emotion prediction.

RQ1:

Overall, we found that the **CNN made emotion and attention predictions with significantly higher accuracy than the analysts** on any given input data. For emotion, we found that the CNN accuracy can be as high as 2.8 times the average accuracy of analyst. Similarly for attention (2.7 times as high).

Fig 5: (Both input) The CNN has more data points due to smaller chunks of time. The ground truth distribution skews towards Excited and Relaxed emotion and the analysts also are biased towards choosing Excited and Relaxed, although incorrectly most of the time.



RQ2:

We found that the back-of-device pressure data impact performance for the CNN model regardless of tasks, but only limitedly for analyst specific to the Dual-stick shooter game. Using pressure data on the CNN, there is an average of 21.2% accuracy increment for emotion prediction and 10.2% for attention prediction, across all tasks. Analysts achieved the highest accuracy in most cases using only pressure input.

The analysts had comparable accuracy using either pressure-only or motion-only input and the lowest accuracy using both input. The CNN had comparable accuracy using both input or using pressure-only input, while having significantly lower accuracy with motion-only data.

Discussion

Effect of Using Non-Contextual Data For Predictions

The analysts heavily confused Excited with Relaxed, suggesting ambiguity in data presented. This might suggest **that both pressure data and orientation might not be apt for human analysts to infer emotions or attention levels** from since those data modalities do not offer behavioral cues that humans largely rely on to form their readings.

Back-of-Device Pressure Pad

Our results show that training ML models on back-of-device pressure data or device's motion can achieve very high accuracy. This can inform researchers about the predictive capability of back-of-device interactions and the use of less obtrusive data modalities in affect prediction

Conclusion

Using motion and back-of-phone pressure, we trained a CNN model which achieves 60%–89% accuracy for predicting one of four emotion states, and 80%–96% accuracy for predicting one of three attention levels. Motion-only data leads to predictions on the lower end of those ranges, while sensing back-of-phone pressure using custom-developed hardware tends to the higher end of the ranges. The predictions by the two human analysts achieved only marginally better than random chance, but the automated model performed model significantly better in a similar task.

Future Work

For many studies, it would be worth investigating how much the additional context of seeing and hearing what the user sees and hears, would be to inferring their emotion and attention.

We envision incorporating the CNN model into the replay and annotation application as an option for automatically inferring the emotion state and attention level of the user during the user session.

[1] Jing Qian et al. Remotion: A Motion-Based Capture and Replay Platform of Mobile Device Interaction for Remote Usability Testing

[2] Enrique Garcia-Ceja, et al. 2015. Automatic stress detection in working environments from smartphones' accelerometer data: a first step

[3] Lucas Paletta, et al. 2014. Attention in mobile interactions: Gaze recovery for large scale studies

[4] Robert LiKamWa et al. 2013. Moodscope: Building a mood sensor from smartphone usage patterns.

[5] Surjya Ghosh et al. 2017. Tapsense: Combining self-report patterns and typing characteristics for smartphone based emotion detection