



# Toadstool: A Dataset for Training Emotional Intelligent Machines Playing Super Mario Bros

Henrik Svoren\*  
henrik.svo@gmail.com  
University of Oslo  
Oslo, Norway

Vajira Thambawita‡  
vajira@simula.no  
SimulaMet  
Oslo, Norway

Pål Halvorsen‡  
paalh@simula.no  
SimulaMet  
Oslo, Norway

Petter Jakobsen†  
peja@helse-bergen.no  
NORMENT, Haukeland University  
Hospital  
Bergen, Norway

Enrique Garcia-Ceja  
enrique.garcia-ceja@sintef.no  
SINTEF Digital  
Oslo, Norway

Farzan Majeed Noori  
farzanmn@ifi.uio.no  
University of Oslo  
Oslo, Norway

Hugo L. Hammer\*  
hugoh@oslomet.no  
OsloMet  
Oslo, Norway

Mathias Lux  
mlux@itec.aau.at  
Alpen-Adria-Universität Klagenfurt  
Klagenfurt, Austria

Michael Alexander Riegler§  
michael@simula.no  
SimulaMet  
Oslo, Norway

Steven Alexander Hicks‡§  
steven@simula.no  
SimulaMet  
Oslo, Norway

## ABSTRACT

Games are often defined as engines of experience, and they are heavily relying on emotions, they arouse in players. In this paper, we present a dataset called *Toadstool* as well as a reproducible methodology to extend on the dataset. The dataset consists of video, sensor, and demographic data collected from ten participants playing *Super Mario Bros*, an iconic and famous video game. The sensor data is collected through an Empatica E4 wristband, which provides high-quality measurements and is graded as a medical device. In addition to the dataset and the methodology for data collection, we present a set of baseline experiments which show that we can use video game frames together with the facial expressions to predict the blood volume pulse of the person playing *Super Mario Bros*. With the dataset and the collection methodology we aim to contribute

to research on emotionally aware machine learning algorithms, focusing on reinforcement learning and multimodal data fusion. We believe that the presented dataset can be interesting for a manifold of researchers to explore exciting new interdisciplinary questions.

## CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Computing methodologies** → *Visual inspection; Neural networks; Classification and regression trees.*

## KEYWORDS

Multimedia Datasets, Neural Networks, Emotional Machines, Machine Learning

### ACM Reference Format:

Henrik Svoren, Vajira Thambawita, Pål Halvorsen, Petter Jakobsen, Enrique Garcia-Ceja, Farzan Majeed Noori, Hugo L. Hammer, Mathias Lux, Michael Alexander Riegler, and Steven Alexander Hicks. 2020. Toadstool: A Dataset for Training Emotional Intelligent Machines Playing Super Mario Bros. In *11th ACM Multimedia Systems Conference (MMSys'20)*, June 8–11, 2020, Istanbul, Turkey. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3339825.3394939>

## 1 INTRODUCTION

"Stop Dave. Stop Dave. I am afraid. I am afraid Dave." This iconic quote from Stanley Kubrick's *2001: A Space Odyssey* is taken from a scene where the sentient computer system HAL 9000 is pleading for life as the human operator is about to shut it down. The movie was released in 1969, and looking at the state of artificial intelligence

\*Also affiliated with SimulaMet, Norway

†Also affiliated with University of Bergen, Norway

‡Also affiliated with Oslo Metropolitan University, Norway

§These authors share the senior position of the article.

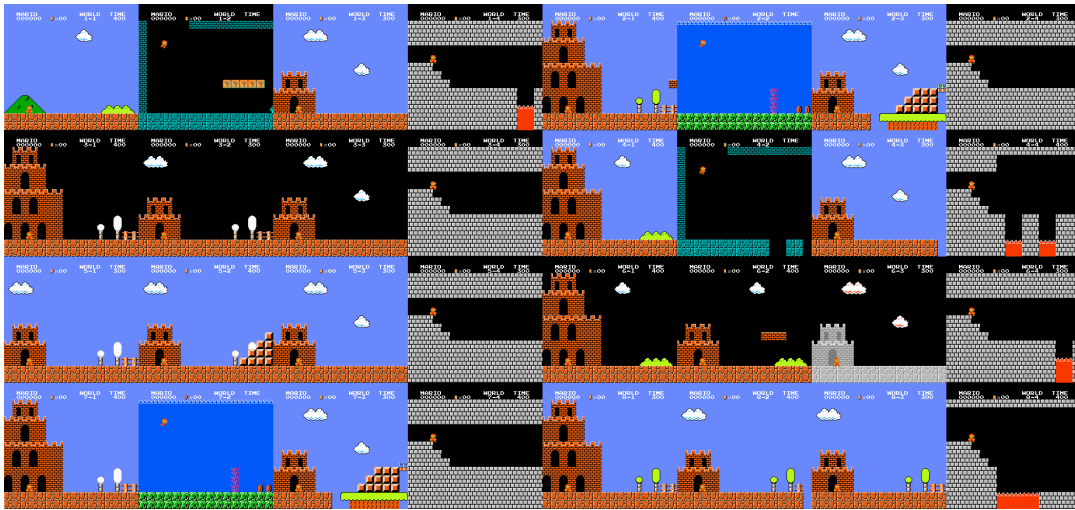
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*MMSys'20*, June 8–11, 2020, Istanbul, Turkey

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6845-2/20/06...\$15.00

<https://doi.org/10.1145/3339825.3394939>



**Figure 1:** Frames are taken from each of the 32 levels contained within *Super Mario Bros.* Note that each image is taken from the very first frame of each level. Levels in *Super Mario Bros.* are organized in groups of four and called worlds, so the first level is world 1-1, the second level is world 1-2, the fifth level is world 2-1, etc.

(AI) today, we can make two observations. First, people in the 60s and 70s were very optimistic about the future capabilities of AI. Second, we are far away from anything near the emotional intelligence that HAL 9000 expresses throughout the movie. For the most part, current AI systems are focused on performing well on specific tasks like classification, object detection or regression, while a machine that can express general intelligence is still far off. This is not negative in and of itself [7], but it is quite different from what people in the past imagined AI would be in the future, and what we might imagine today.

Using machine learning to interpret or detect human emotions is a growing field of research. This is commonly done using different types of media, such as images [2], sensor data [11], text [23], or some combination of the three [4, 14]. Recent works in this field have also moved to look at how human emotion data may affect the training and performance of deep learning algorithms. McDuff et al. [17] explore how human emotional response may affect the performance of a self-driving agent trained in a simulated environment. They showed that adding human-like signals, such as the blood volume pulse (BVP), helped improve the driving performance of the algorithm. The idea of supplementing today's machines with emotional or physiological signals is supported by the large amount of literature that shows that pure rational decision making is often not optimal in humans [3, 9, 18, 22]. Prior research shows that emotional content can help guide the decision-making process as well as make it more efficient [16]. Some early work also tried to use this for artificial agents [10]. Such findings suggest the possibility that similar benefits might be had by artificial agents, especially when engaged in human-like tasks or behavior.

Inspired by the work done by McDuff et al. [17], we look at other areas where the same principles may be applied, which in this case, is playing the well-known classical video game *Super Mario Bros.* While this game is not representative for all video games that are available right now, it is commonly accepted as a well-known, good example for a video game and can be considered representative

for the jump and run and the arcade game genres. To perform experiments in that direction, we first need a dataset that contains both the frames from *Super Mario Bros.* and the sensory output of the player. As no such dataset exists, we collected gameplay data, sensor data, and facial expression data from ten different participants. Furthermore, we also made the dataset and all sources to re-produce the games played publicly available. We think this dataset is of great interest to many research communities as it consists of multiple modalities and is applied to a unique use case. The contributions of this paper are three-fold:

- (1) We present a publicly available, multimodal dataset which focuses on the human component of intelligent machines along with a reproducible methodology to extend the dataset with additional data collection.
- (2) We present a set of baseline experiments that aim to show how the dataset can be used to predict specific sensor values using a combination of data from the video game and facial expressions.
- (3) We outline future applications and interesting research questions using the dataset.

To the best of our knowledge, this is the first open dataset that provides the (i) video frames of a person's facial expressions, (ii) the sensory output of the person playing a game, and (iii) data from the video game synchronized with the facial expressions and sensor data. The dataset opens up for a wide range of new and interesting analyses, and a proper and fair comparison between different methods, both from a psychological and a multimedia perspective. In the following, the process of collecting the data, as well as the resulting data, are described. Moreover, a baseline evaluation is presented, including suggestions for future research directions using the dataset.

ID	Age	Sex	Dominant hand	Hours per week	Years active	Prior experience	Game score
0	26	Male	Right	4-8	22	Lots	17, 100
1	48	Male	Left	0-1	1	Little	3, 000
2	28	Male	Right	0-1	0	None	300
3	32	Male	Right	4-8	4	Some	13, 300
4	32	Female	Right	0-1	5	Some	6, 400
5	30	Female	Right	0-1	5	Little	2, 700
6	35	Male	Left	1-4	30	Lots	14, 300
7	34	Female	Right	1-4	14	Some	3, 800
8	31	Female	Right	0-1	2	Little	200
9	27	Female	Right	0-1	5	Little	10, 600

**Table 1: This table shows an overview of all participants included in the dataset.**

## 2 DATA COLLECTION

The dataset was collected at our research laboratory located in Oslo, Norway. Participants were selected based on a set of criteria, mostly focused on their prior gaming experience. We wanted to collect data from people with a wide range of different game experience backgrounds. This includes those who have barely touched a video game to those who have been playing since childhood. Furthermore, we aimed to collect data from a balanced set of genders, meaning an even split of male and female participants. Each participant was asked to fill out a short questionnaire about their previous video game experience in addition to some information about themselves. An overview of the answers can be seen in Table 1. In total, ten participants were selected for the study, where each participant provided a written form of consent, allowing for their video, gameplay data, and sensor data to be shared openly for research and teaching purposes under the license Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)<sup>1</sup>. The dataset can be accessed via (<https://datasets.simula.no/toadstool>) or (<https://osf.io/qrkcf/>).

As for collecting the gameplay data, we developed a protocol that describes what data should be collected and how. This protocol went through multiple iterations as we performed a preliminary test run before applying it to all participants in the study. From this initial test run, we learned that, in some cases, the conductivity between the participant and the wristband (Empatica E4) did not gather data in line with what we expected. Some anomalies included little to no detected activity and substantial value differences between participants. Furthermore, we noticed that the activity would vary a lot between the start and end of a gameplay session. The primary cause of this was mostly due to the dry conditions in which the data was collected. For the wristband to accurately pick up a person’s sensor data, the electrodes need some sweat to act as a conductor between the skin and wristband. Based on these observations, we changed the protocol to include a short warm-up session before playing the video game and a 15-minute period where the participant would sit still to develop a baseline. The warm-up consisted of walking up and down a flight of stairs spanning six floors two times. This exercise was selected based on tests with some people in the laboratory. The final protocol is shared with the dataset.

Before playing, the participants were told that their performance in the game would be measured based on how many stages were cleared in the time given and on the number of player avatar

deaths. Furthermore, we informed participants that their performance would be measured against other participants and that there would be a prize for the highest achiever. The motivation behind making the game more competitive was to make the players want to perform well, and feel like there was some consequence if they either died (in the game) or did not beat levels fast enough. The number of points earned by each player was kept secret from all participants to avoid them giving up or relaxing due to other players too high or too low score. Scores were calculated based on two primary factors; the number of deaths and levels cleared. Starting a level, the player starts with a base score of 1000. For every death, the score is reduced by 100 points down to a minimum of 200. If the player manages to beat the level, he/she is awarded between 200 and 1000. If the player runs out of time, he/she is awarded 0 points and is moved to the next stage. The final score of each participant is included in the dataset and can be seen in Table 1.

After the participants had established a sufficient baseline, they started the primary game session where the video of the participants, video game frames, and sensor data was collected. Each game session lasted for approximately 35 minutes. The game was played directly in the gym-super-mario-bros environment [13], which is a gym [1] based environment for *Super Mario Bros*. For reproducibilities sake, the repository for the gym environment has been added to the official GitHub repository of Toadstool<sup>2</sup>. There are four different graphic environments offered by gym-super-mario-bros, which include the standard graphics, as well as three different downsampled versions (downsample, pixel, and rectangle). The data in our dataset is collected in the standard environment, but sessions may be replayed in any environment if needed.

The gym environment version of the game still differs somewhat from the original gaming experience found in *Super Mario Bros*. by Nintendo (a consumer electronics company from Japan). Firstly, all game-freezing animations and cutscenes are removed from the game. This includes transitions between levels and traveling through pipes. Second, there is no music or other sound effects. Third, there are no limits on game lives, and power-ups do not carry over to new stages. Last, the order of the levels has been changed compared to the original game. There was one additional rule we told participants before playing. In *Super Mario Bros.*, some pipes can warp the player to a new stage that is closer to the final stage of the game. To keep the levels played consistent between players,

<sup>1</sup><https://creativecommons.org/licenses/by-nc/4.0/>

<sup>2</sup><https://github.com/simula/toadstool>

we asked participants to refrain from using any of the available warp pipes (one located in world 1-2 and two located in world 4-2). Overall, it took approximately one hour to collect data from a single participant.

### 3 DATASET DETAILS

For each participant, we have included a video of them playing the game (camera facing the face), the controller input performed on each frame of the game, and the sensor data collected from an Empatica E4 wristband [6]. The camera used to collect the facial expression data was a 1.3-MP webcam attached to a Samsung Series 9 Notebook NP900X4C. The webcam captured video at 30 frames per second with a resolution of  $640 \times 480$ . The controller used to play the game was a wired USB controller from retro-bit, which is modeled after the original controller for the Nintendo Entertainment System. Note that the video game frames are not included in the dataset, but can be extracted by using the provided video game actions files included with each participant. This can be done by using a script that is included in the dataset. The reason for not including the video game frames was mostly due to the exponential increase in storage size. Another possible benefit of this approach is the ability to replay the game session in any of the several environments offered by the gym-super-mario-bros framework to produce different representations of game frames. The first frame for each of the 32 levels of *Super Mario Bros.* can be seen in Figure 1. The dataset contains the following files:

- **participants** is the directory that contains the information of each participant. This includes the video of them playing, the controller input of each game frame, and the Empatica E4 wristband sensor data.
- **scripts** is a directory that holds a set of Python scripts meant to aid the user in getting an easy start to using the dataset. The files include a script for replaying gameplay using the provided controller inputs, a script for matching the gameplay session to the facial expression video, and a script for matching the raw signal outputs to the gameplay session.
- **protocol.pdf** is the protocol used to collect the video game session data.
- **questionnaire.pdf** is the questionnaire that was filled out by each participant before starting the game session.
- **questionnaire\_answers.csv** is a summary of all the answers to the questionnaire.
- **consent.pdf** is the consent form that was signed by each participant.
- **README.txt** is a short information file which describes the contents of the dataset.
- **LICENSE** is the file that signifies which license in which the dataset is distributed under.

Contained within the *participants* directory is a separate directory per participant included in the dataset. Each directory has a name corresponding to the participant's ID, i.e., *participant\_<ID>*, where *<ID>* is replaced with the ID of the participant. For each participant, we have stored the participant's sensor data collected from the Empatica E4 wristband, a JSON file containing information about the participants game session, the video recording of the participant playing the game stored in ".avi" format, and another JSON

file which contains information about the video. The JSON file that holds the game session data, called *participant\_<ID>\_session.json*, contains the actions performed during the game, the start and end times of the game session, and the achieved gameplay score. As for the sensor data, the wristband uses four separate sensors to collect different sensory outputs from the wearer, such as the electrodermal activity (EDA), interbeat intervals (IBI), heart rate (HR), and blood volume pulse (BVP). The four sensors of the wristband is a photoplethysmography sensor, an electrodermal activity sensor, 3-axis accelerometer, and an optical thermometer. Of the four sensors, the thermometer is the only one not graded for clinical use. All data collected by the wristband are stored in CSV files that can be downloaded from the wristband. For the dataset, these CSV files have been matched to the game session. We have also opted to include the raw source files as they were collected from the wristband. The CSV files and a short description of the contents are further explained below.

- **ACC.csv** contains the data collected from the 3-axis accelerometer sensor in the range  $[-2g, 2g]$  sampled at 32 Hz. The accelerometer measures the movement of the wearer.
- **EDA.csv** holds the data collected by the EDA sensor sampled at 4 Hz. EDA measures the electrical conductivity of the skin and measurements have been proven to be correlated with emotions since the late 1800s [5]. EDA is also sometimes called psychogalvanic reflex or skin conductance.
- **BVP.csv** contains the data collected from the photoplethysmograph sensor, which measures the BVP, and is sampled at 64 Hz.
- **IBI.csv** stores the interbeat intervals (IBI). The IBI measures the time interval between individual heartbeats and can be used to estimate the instantaneous heart rate as well as heart rate variability. The wristband calculates the values contained within this file based on the BVP signals.
- **HR.csv** contains the average heart rate values, computed in spans of 10 seconds. The heart rate measures the number of times a person's heartbeats per minute. Similar to the IBI, these values are calculated based on the BVP signal.
- **TEMP.csv** holds the information collected by the thermometer, which is the temperature of the person playing the game expressed in degrees Celsius ( $^{\circ}\text{C}$ ) sampled at 4Hz.
- **info.txt** gives a brief description of all variables collected by the wristband.

In addition to the data collected throughout the study, we also include a set of scripts that aim to make the dataset more accessible. First of all, as previously mentioned, video game frames are not included in the dataset. However, we include the necessary information to extract the frames by using the provided video game inputs to "replay" the game session and collect the video frames directly.

### 4 PRELIMINARY EXPERIMENTS

We performed a set of preliminary experiments to showcase how the presented dataset can be used to train machine learning algorithms and perform simple predictive modeling. In Section 5, we mention that a possible use case for the dataset is to predict the sensor value of the wristband using the video game frames or facial

ID	CNN		ZeroR	
	MAE	RMSE	MAE	RMSE
0	0.076	0.100	0.075	0.099
1	0.104	0.132	0.103	0.131
2	0.071	0.104	0.075	0.100
3	0.050	0.070	0.050	0.070
4	0.078	0.103	0.069	0.094
5	0.091	0.129	0.094	0.121
6	0.091	0.119	0.090	0.116
7	0.110	0.142	0.109	0.139
8	0.061	0.096	0.060	0.090
9	0.126	0.157	0.109	0.134
All	0.105	0.126	0.090	0.117

**Table 2: This table shows the results of all experiments of trying to predict the BVP amplitude using video game frames and facial expressions alone. Note that all experiments were trained over three-fold cross-validation.**

expressions as input. In the following experiments, we trained a deep convolutional neural network to predict the BVP amplitudes utilizing a combination of the face data and game frame data. This is similar to what McDuff et al. [17] did when modeling the *emotional* input to their self-driving reinforcement agent.

Before the training step, we needed to prepare the input data, i.e., the video game recording and facial expression video. As the gameplay frames were recorded at 60 frames per second, while the facial expression video was recorded at 30 frames per second, we down-sampled the video game frames by skipping every other frame. After that, the BVP amplitudes had to be extracted from the raw BVP signals and matched to the input frames. The reason for not predicting the raw BVP values is due to the cyclic nature of a beating heart. The BVP has the properties of a sine-like wave that is composed of valleys and peaks which appear in tandem with every heartbeat. We are not interested in the exact value on the signal curve, but the highest point of a cardiac cycle, also known as the systolic peak. This peak-value gives us some information about the emotional state of the human player [20]. To extract the BVP amplitudes, we detect a systolic peak in the given BVP signal and measure its height from the baseline. This peak-value is then repeated until the next detected peak, and-so-forth. The result is a square-like signal in comparison to the sine-like wave that is the BVP. The BVP values were then matched with the input data by taking the average BVP amplitude values over one second (64 measurements in total). The extracted peaks used for the experiments are shared on GitHub together with the code used to produce the baseline experiments<sup>3</sup>.

We trained one convolutional neural network (CNN) per participant in addition to one trained on all participants mixed. The purpose of training a model on a single participant is because we generally want to model the emotional response of a single person, not necessarily everyone at once. The model was based on the TensorFlow implementation of ResNet50 [8], where we input two video game frames (first and the last frame of one second) and the

facial expression from the last frame of the second corresponding to the video game frames. These three images were grayscale and stacked channel-wise before being processed by the model. In addition to the CNN-based model, we also calculated the error when using the mean of the response variable in the training for prediction (also sometimes called ZeroR), which should give us some indication about how well our model performs. Since the response is a continuous measurement, we used the metrics mean absolute error (MAE) and root mean squared error (RMSE). Furthermore, we used three-fold cross-validation to not bias the results towards a pre-defined split of the data.

The results of predicting the BVP amplitudes using video game frames and facial expressions can be seen in Table 2. We observe that the trained model sometimes outperforms the baseline and sometimes not, which is an indicator of a challenging task, but not impossible. There is still much room for improvement, where methods recurrent neural networks may be used to increase performance, but this is out of scope for this paper. Furthermore, we want to point out that for classification tasks, such as classifying emotional states based on sensor and video data, one should use standard classification metrics such as precision, recall/sensitivity, and F1-score to determine the quality of the trained model. Optimally, one should also report the true positives, true negatives, false positives, and false negatives so that readers themselves can calculate the metrics manually. The code used to run all experiments is available online in the datasets official GitHub repository (<https://github.com/simula/toadstool>).

## 5 POSSIBLE APPLICATIONS OF THE DATASET

We expect that this dataset can be used for many different use cases and scenarios. First of all, we imagine it can be used to detect relationships between the player sentiment and the current gameplay state. This could solely be based on the gameplay frames and collected sensor data, or could also be combined with the player’s facial expression for further analysis. The uncovered relationships could be interesting when studying how players get invested in video games, and what typical scenarios contribute to a strong reaction from players. A more specific example could be predicting a person’s facial expression based on a given game state or degree of progress in a game or level. This could also be expanded to the sensor data, as one could predict the sensory output of the wristband using the game state and facial expressions as input, like photoplethysmography [21] does with the heart rate, but connected to the current game state the player is in. These two problems could be modeled as either a regression or classification problem, depending on the application.

From a game design perspective, correlations between game progress, game state, the input of players, and the emotions and sentiment of the players could enable a whole new approach to experience-based game design. With the possibility to use this as a method for playtesting, game designers can evaluate when and how their crafted experience is (or is not) invoked in the players. Moreover, games can be built around the concept of self-adapting challenge and difficulty by counter-acting unnecessary frustration, boredom, or annoyance by adapting the game to the player’s emotion and sentiment. Last but not least, the correlation between the

<sup>3</sup><https://github.com/simula/toadstool>

game, emotions, sentiment, and game engagement, especially flow (the state of being fully immersed in an activity while enjoying it), can be investigated [15, 24]. This would lead to a better understanding of what makes games enjoyable and would impact fields like game and media studies, psychology, game engineering, game design, and serious games / educational games.

Another area where this dataset could be used is in the training of emotionally intelligent machines using reinforcement learning. One way to do this would be by training an algorithm to reproduce the physiological signals based on the game-state. The reproduced signals can then be incorporated into the reward function of a reinforcement learning agent. Since physiological signals like BVP and EDA are reliable indicators of emotional states in humans [12, 19], such an approach could be used to mimic human emotional responses. These kinds of emotionally intelligent machines might aid the pure logic of reinforcement learning models and produce improved learning, as well as reveal new insights into how both machines and humans learn. They might also be a step towards machines with even more complex emotional intelligence, like the ability to recognize, express, and respond to human emotions. This is an active research area leading to better personal assistants, more believable automated communication, as well as more enjoyable and believable video game AI.

Some more concrete examples of possible research questions or experiments are:

- How are the sensor measurements related to the facial expressions?
- Can sentiment analysis of the facial expression be connected to measurements in the sensors?
- How can different data sources be combined efficiently (sensor data with videos, etc.)?
- Can game actions of the human players be used as a baseline for human performance?
- Can the additional data collected from people be used to train reinforcement learning algorithms?
- Would a model trained on input from non-experienced players behave differently from a model trained on experienced players?
- Can immersion, enjoyment, and flow automatically be inferred from the gathered data?

As one can see from the discussion above, the dataset holds a lot of interesting research potential to follow up, and we hope that other researchers get inspired to work on the dataset.

## 6 CONCLUSION

In this paper, we present *Toadstool*, a new dataset consisting of people playing *Super Mario Bros.* and physiological signals corresponding to their emotional reaction to playing said game. While the dataset provides a good starting point for applied research in affective computing in games, emotional AI agents in games, playtesting, and game analytics, we strongly believe the *Toadstool* dataset will foster research in many ways and, therefore, we have detailed follow up research questions in several fields, including affective computing, psychology, and game and media studies. With the detailed description of the dataset, the in-depth discussion of the method included in the dataset, and the wide availability of the

game instance used for the experiment and the medical sensors employed, we ensured reproducibility and extension of the dataset. We hope that our work inspires and encourages interdisciplinary and multidisciplinary research to (i) examine how the human element in human-computer interaction can be employed to improve existing machine learning methods by introducing emotional aspects to an otherwise cold and unfeeling machine and (ii) show how interactive entertainment systems, and especially video games, utilize sensory input to provide a personalized and tailored user experience.

## REFERENCES

- [1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. arXiv:arXiv:1606.01540
- [2] Prudhvi Raj Dachapally. 2017. Facial emotion detection using convolutional neural networks and representational autoencoder units. *arXiv preprint arXiv:1706.01509* (2017).
- [3] Benedetto De Martino, Dharshan Kumaran, Ben Seymour, and Raymond J Dolan. 2006. Frames, biases, and rational decision-making in the human brain. *Science* 313, 5787 (2006), 684–687.
- [4] Liyanage C De Silva, Tsutomu Miyasato, and Ryohei Nakatsu. 1997. Facial emotion recognition using multi-modal information. In *ICICS 1997*, Vol. 1. IEEE, 397–401.
- [5] Ch Fere. 1888. Note sur des modifications de la resistance electrique sous l'influence des excitations sensorielles et des motions. *Compt Rend Soc de Biol* 8 (1888), 217–219.
- [6] M. Garbarino, M. Lai, D. Bender, R.W. Picard, and S. Tognetti. 2014. Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *ICWMCHM 2014*. 39–42.
- [7] Uri Hasson, Samuel A Nastase, and Ariel Goldstein. 2020. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron* 105, 3 (2020), 416–434.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *IEEE CVPR*. 770–778.
- [9] Michael Hechter and Satoshi Kanazawa. 1997. Sociological rational choice theory. *Annual review of sociology* 23, 1 (1997), 191–214.
- [10] Hong Jiang and Jose M Vidal. 2006. From rational to emotional agents. In *AAAI CMASS 2006*.
- [11] Eiman Kanjo, Eman M.G. Younis, and Chee Siang Ang. 2019. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion* 49 (2019), 46 – 56.
- [12] Krisztian Kasos, Szabolcs Zimonyi, Eniko Kasos, Avraham Lifshitz, Katalin Varga, and Anna Szekely. 2018. Does the electrodermal system “take sides” when it comes to emotions? *APBF Journal* 43, 3 (2018), 203–210.
- [13] Christian Kauten. 2018. Super Mario Bros for OpenAI Gym. <https://github.com/Kautenja/gym-super-mario-bros>.
- [14] Anil Kumar K.M., Kiran B.R., Shreyas B.R., and Sylvester J. Victor. 2015. A Multimodal Approach To Detect User's Emotion. *Procedia Computer Science* 70 (2015), 296 – 303.
- [15] Derek A Laffan, John Greaney, Hannah Barton, and Linda K Kaye. 2016. The relationships between the structural video game characteristics, video game engagement and happiness among individuals who play video games. *Computers in Human Behavior* 65 (2016), 544–549.
- [16] Jennifer S. Lerner, Ye Li, Piercarlo Valdesolo, and Karim S. Kassam. 2015. Emotion and Decision Making. *Annual Review of Psychology* 66, 1 (2015), 799–823.
- [17] Daniel McDuff and Ashish Kapoor. 2019. Visceral Machines: Risk-Aversion in Reinforcement Learning with Intrinsic Physiological Rewards. In *ICLR 2019*.
- [18] David Palumbo-Liu. 2005. Rational and Irrational Choices: Form, Affect, and Ethics. *Minor Transnationalism* (2005), 41–72.
- [19] Rosalind W. Picard, Szymon Fedor, and Yadid Ayzenberg. 2016. Multiple Arousal Theory and Daily-Life Electrodermal Activity Asymmetry. *Emotion Review* 8, 1 (2016), 62–75.
- [20] R. W. Picard, E. Vyzas, and J. Healey. 2001. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on PAMI* 23, 10 (Oct 2001), 1175–1191.
- [21] Philipp V Rouast, Marc TP Adam, Raymond Chiong, David Cornforth, and Ewa Lux. 2018. Remote heart rate measurement using low-cost RGB face video: a technical literature review. *Frontiers of Computer Science* 12, 5 (2018), 858–872.
- [22] John Scott. 2000. Rational choice theory. *Understanding contemporary society: Theories of the present* 129 (2000), 671–85.
- [23] Shiv Naresh Shivhare and Saritha Khethawat. 2012. Emotion detection from text. *arXiv preprint arXiv:1205.4944* (2012).
- [24] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)* 3, 3 (2005), 3–3.