

Collection and Validation of Psychophysiological Data from Professional and Amateur Players: a Multimodal eSports Dataset

¹Anton Smerdov, ²Bo Zhou, ²Paul Lukowicz, ¹Andrey Somov

¹Skolkovo Institute of Science and Technology, CDISE, Moscow, Russia

²German Research Centre for Artificial Intelligence, Kaiserslautern, Germany

arXiv:2011.00958v2 [cs.HC] 23 Aug 2021

Abstract—Proper training and analytics in eSports require accurately collected and annotated data. Most eSports research focuses exclusively on in-game data analysis, and there is a lack of prior work involving eSports athletes’ psychophysiological data. In this paper, we present a dataset collected from professional and amateur teams in 22 matches in League of Legends video game with more than 40 hours of recordings. Recorded data include the players’ physiological activity, e.g. movements, pulse, saccades, obtained from various sensors, self-reported after-match survey, and in-game data. An important feature of the dataset is simultaneous data collection from five players, which facilitates the analysis of sensor data on a team level. Upon the collection of dataset we carried out its validation. In particular, we demonstrate that stress and concentration levels for professional players are less correlated, meaning more independent playstyle. Also, we show that the absence of team communication does not affect the professional players as much as amateur ones. To investigate other possible use cases of the dataset, we have trained classical machine learning algorithms for skill prediction and player re-identification using 3-minute sessions of sensor data. Best models achieved 0.856 and 0.521 (0.10 for a chance level) accuracy scores on a validation set for skill prediction and player re-id problems, respectively. The dataset is available at https://github.com/smerdov/eSports_Sensors_Dataset.

Index Terms—eSports, dataset, machine learning, psychophysiological assessment, sensing, video games

I. INTRODUCTION

Competitive video gaming, or eSports, has gained tremendous popularity within the last several years. eSports has evolved into a mature industry with well-funded tournaments, professional athletes, and a vast fan community. The strengthened competition requires professional eSports teams to explore new methods for training and analytics. In its turn, it drives the demand for eSports research.

A common and natural source of data for eSports research is in-game data. The majority of prior research relies on information obtained from the game logs, e.g. heroes drafted, abilities learned, players’ positioning [1]–[3]. This information may be helpful in evaluating the players’ performance, predicting the outcome of the match, understanding players’ behavior, and other analytics. However, if only in-game data are used, much information from the real world is omitted. Sensor data obtained from the physical domain can supplement in-game logs collected in the digital domain [4]–[6]. Similarly, predictive models trained on sensor data can supplement regular

models trained using in-game data and provide additional evidence for analytics [7], [8]. Data obtained from sensors can represent players’ physical conditions, psychological state, environmental situation, and other factors that may help in eSports analytics. Investigation of the connection between the physiological data and player performance may help provide personalized feedback for efficient training.

The major disadvantage of using only in-game data is the poor robustness of predictive models. These models capture dependencies between the in-game parameters and some target values. When rules or game mechanics change, previous dependencies utilized to fit the model might not persist. Given the regularity of patches and updates in eSports games versions, a tight connection between the in-game parameters and model prediction can significantly limit model usage lifespan. It makes the models based on real-world data collected from sensors practically feasible since they are not sensitive to game mechanics changes.

Although the idea to use sensor data in eSports research is not new, prior work on this matter is typically limited to one sensor used or one player recorded [9]–[11]. That narrows the insights explored as the multi-sensor dependencies are not captured, and person-to-person interactions are not registered. Recording with multiple sensors might help understand which signal is more important in the eSports domain and improve the model performance and robustness since the data from multiple sources are utilized. Capturing sensor data from multiple players simultaneously can help capture interactions between players and investigate team dynamics, which is especially relevant because many eSports games are team games.

In this work, we present a dataset of sensor data collected from 10 participants playing League of Legends, 22 matches in total. Participants were invited based on their skill level and organized in two teams of 5 players: professional players with vast experience in the game and amateur players with basic experience in League of Legends. The sensor data collected include information on hand/head/chair movements, heart rate, muscle activity, gaze movement on the monitor, galvanic skin response, electroencephalography (EEG), mouse and keyboard activity, facial skin temperature, and environmental data. Additionally, we provide feedback from each player after each match with their feeling about each match.

Every match is accompanied by a game log indicating key in-game events like kills/deaths/assists, players positioning, skills learned, buildings destroyed, etc. The dataset is structured as directories for each of 22 matches, and each match consists of directories for each of 5 players with many .csv files with sensor data presented.

We consider the extreme setting when only sensor data are available for prediction, and in-game data are used for labeling only. It allows us to train the model that is independent from in-game parameters and check if sensor data are enough to properly understand the players' behavior.

Contributions of this research is twofold:

- 1) An extensive dataset with sensor data collected from players with various skill levels.
- 2) A team dynamics study in the eSports domain to show the validity of data collected.

In terms of novelty, we carry out the data collection from *professional* players (apart from the amateur ones) recording the data from five players at a time. The collected data are truly multi-modal including the physiological data, in-game data, and self-reported survey.

This paper is organized as follows: in Section II we overview the state-of-the-art methods using sensor data for performing the data analysis in eSports; in Section III we present a sensor network architecture for data collection; in Section IV we describe the methodology of data collection; in Section V we describe in details the data collected; in Section VI we provide concluding remarks.

II. RELATED WORK

State-of-the-art work in eSports research can be divided into two categories: (i) a wide group of studies exploring dependencies in in-game data, and (ii) a small scope of work investigating the importance of psychophysiological signals in the eSports domain. In this section we review methods utilizing sensor data in eSports and discuss competing approaches which use data only from the digital domain.

A. Psychophysiological Research in eSports

Although psychophysiological research in eSports is still in its infancy, there is a number of prior studies utilizing one or a few sensors for data collection.

A popular sensor for eSports research is an eye tracker. Using the data obtained from an eye tracker, authors in [11] suggest that the crucial trait of professional players might be the duration and variability of visual fixations. This idea is elaborated further in [7] where authors propose to classify players into professional and amateurs based on their reaction time. They calculate the whole reaction time as a sum of three components: saccadic latency, the time between saccade and fixation, and the time for aiming and shooting. In other studies relying on eye tracking for eSports [12] [4] the authors investigate the eye movement patterns able to distinguish the professional players from amateurs. According to their results, professional players' eye-movement patterns are more diverse and swifter.

Another data source in eSports research is EEG. Minchev et al. [9] showed the decrease of EEG alpha rhythms power spectra frequencies for 'lose' game events and increase for the 'win' game events in a first-person shooter (FPS) game; the opposite was shown for the theta rhythm. Another work on EEG in eSports [13] claims that the EEG signal for newbies has more variations than one for experienced players, although only two participants are compared.

Players' movements data can also help to investigate their behavior. In [8] and [14] authors showed that data about chair movements can help to separate high-skilled and low-skilled players in Counter-Strike: Global Offensive. They claim that professional players move on the chair less often but more intensively.

Game input pattern obtained from mouse or keyboard is also used for eSports research. In [10], the authors showed that mouse input data might be useful for player re-identification. Work by Khromov et al. [5] investigated the connection between player's skill and data from mouse, keyboard, and eye tracker. Data obtained from the keyboard turned out to be the most important. Work by Blom et al. [6] presented a dataset with mouse, keyboard, and galvanic skin response for League of Legends players but did not show any dependencies in the data. In another work [15] dedicated to League of Legends, authors presented a dataset with stream videos annotated with streamer affect and game context. They managed to predict the streamer's arousal and game context using convolutional neural networks.

The above mentioned research demonstrate that sensor data can be used in a wide range of eSports problems, such as skill prediction, player re-identification, performance evaluation, etc.

B. In-game Data in eSports

Most of the prior eSports studies rely exclusively on the in-game data collection and further analysis. The straightforward approach is to use in-game indicators such as kills, deaths, and gold to predict match outcome in Multiplayer Online Battle Arena (MOBA) [1], [3], [16]–[18] or FPS genres [19]. More advanced methods also try to mine logic from a predictive model [20], [21] for better interpretation, to extract high-level features (i.e., encounters statistics) from game replays [22], or to calibrate the confidence of prediction [23].

Another approach for match outcome prediction is to use information about the hero drafts to predict a winner before the actual game starts [2], [24]–[26]. These methods can be easily transformed into draft recommendation systems with high practical use.

Genre-agnostic methods for eSports analytics usually rely on match history data and utilize only match results instead of ad-hoc features for a specific game. In [27], the authors leverage players' match history data to estimate players' current psychological state and use this estimation to predict the outcome of matches. Other history-based outcome prediction approaches include machine learning on hand-crafted features for the MOBA genre [28] and modeling with rating systems for MOBA and shooter genres [29]. Work by Dehpanah et

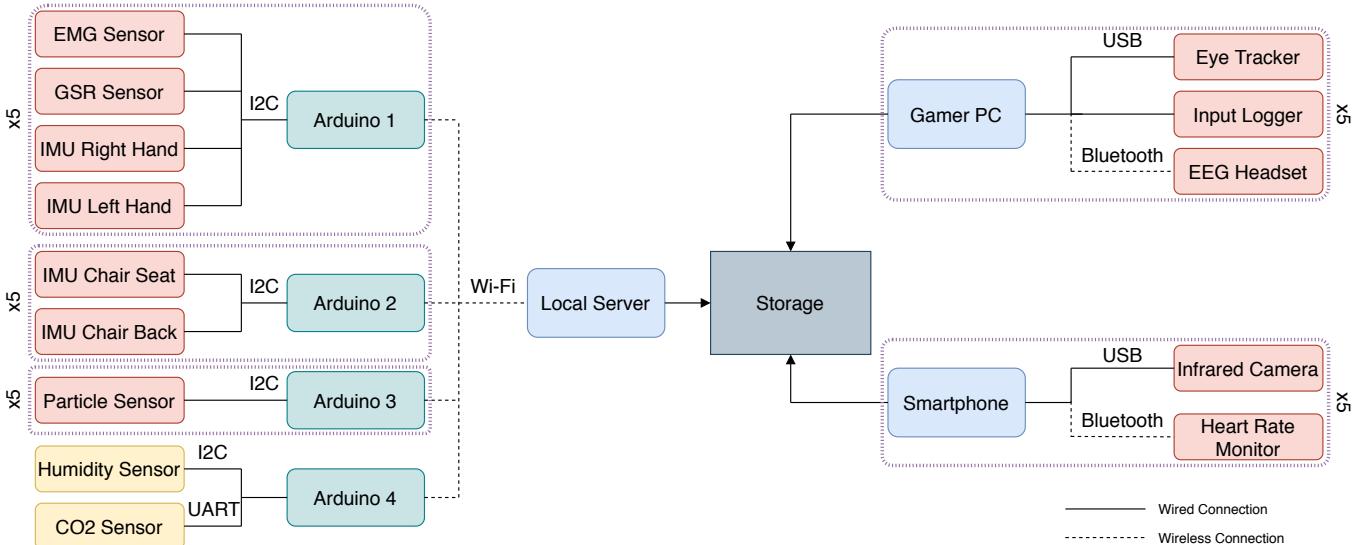


Fig. 1: Sensor network architecture. Solid lines are wired connections; dashed lines are wireless connections.

al. [30] extends rating systems for free-for-all games (Battle Royale genre) using a large dataset of matches in PlayerUnknown's Battlegrounds game.

Analytics in eSports is not limited to winner prediction but also extends to players' behavioral analysis. Authors in [31] proposed to cluster players' in-game data in Massively multiplayer online role-playing games(MMORPG) and strategy games to interpret their behavior and to compose the optimal team lineup. Research in [32] addresses the problem of player re-identification in the MOBA genre. The authors showed that classical machine learning algorithms can predict hero ID and one of three roles based on game data. Pfau et al. [33] emphasized the problem of players' Internet disconnection and proposed to model the behavior of disconnected players with deep learning algorithms. In [34], authors address the problem of toxic behaviors in the eSports games and argue that despite some effort from game developers, this problem persists in eSports community. In [35], authors trying to identify toxic behavior automatically with a toxicity detection algorithm based on the analysis of game chat logs.

III. SENSOR NETWORK ARCHITECTURE

We used twelve types of sensors for data collection:

- 1) Electromyography (EMG) sensor *Grove EMG Sensor v1.1*¹. This sensor measures the intensity of muscle activity and outputs it as voltage (analog interface). It uses ground, plus, and minus electrodes located on a forearm muscle to measure the voltage. The voltage can be interpreted as the intensity of muscle constraint; thus, the intensity of keyboard or mouse strokes or movement. EMG signal is related to physical tension affecting the player's current state [36].
- 2) Galvanic skin response (GSR) sensor *Grove GSR Sensor v1.2*². This sensor outputs voltage (analog interface),

which allows us to infer skin resistance for the participant. Skin resistance is lower in the arousal state due to more sweat, so these data can be used as a stress indicator [37]. Participants placed two electrodes of this sensor on two fingers on the right hand.

- 3) Inertial measurement unit (IMU) *Bosch BNO055*³ located on the wrists on both hands. The sensor recorded linear acceleration, gravity, angular velocity, Euler angles, and quaternions.
- 4) IMU's under the chair seat and on the chair back. Precise locations are shown in Figure 2b. Behavior on a chair is connected with the player skill [8] [14]. Sensors' models and data recorded are the same as from the IMU's located on hands.
- 5) Environmental atmosphere sensor *Bosch BME280*⁴ located in the gaming room. It measures relative humidity and environmental temperature. These data affect players performance: high level of relative humidity results in the decrease of neurobehavioral performance [38]; high environmental temperature may affect human performance [39].
- 6) CO₂ sensor *MH-Z19B*⁵. This sensor measures carbon dioxide concentration in ppm. High CO₂ level may deplete cognitive abilities [40].
- 7) Eye tracker *Tobii 4C* located under the monitor. Data from eye tracker indicate which information on the game screen players check more often, so the effectiveness of their game decisions [11]. To standardize the information obtained with eye trackers, all players ran games in 1920x1080 resolution. Before each experimental day, players calibrated the eye trackers.

¹https://wiki.seedstudio.com/Grove-EMG_Detector/

²https://wiki.seedstudio.com/Grove-GSR_Sensor/

³<https://www.bosch-sensortec.com/products/smart-sensors/bno055.html>

⁴<https://www.bosch-sensortec.com/products/environmental-sensors/humidity-sensors-bme280/>

⁵https://www.winsen-sensor.com/d/files/infrared-gas-sensor/mh-z19b-co2-ver1_0.pdf

- 8) Electroencephalography (EEG) headset *Emotiv Insight*⁶. EEG data have been shown to be connected with arousal [41], so the player performance.
- 9) Input (keyboard and mouse) logger as a python script running on the gamer's PC. These data are indirect indicators of the hand movement activity and the player skill [42].
- 10) Infrared camera *Flir One*⁷ directed to players' faces for facial skin temperature data collection.
- 11) Heart rate monitor *Polar OH1*⁸ armband. High heart rate corresponds to mental stress and arousal [43], which might affect player decisions' rationality.
- 12) Pulse-oximeter sensor *Maxim MAX30102*⁹ located on the right earlobe. Prior research in [44] demonstrated the connection between oxygen saturation and cognitive performance.

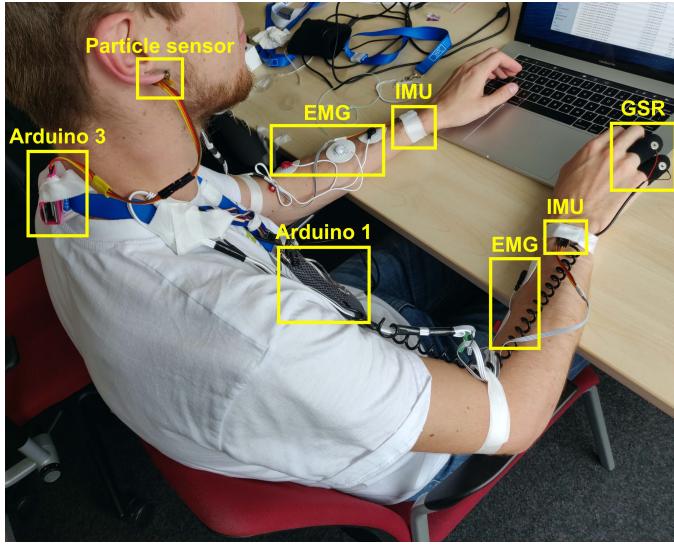
Figure 1 illustrates the sensor network architecture. The experimental testbed is presented in Figures 2a and 2b. We

⁶<https://www.emotiv.com/insight/>

⁷<https://www.flir.eu/flir-one/>

⁸<https://www.polar.com/en/products/accessories/oh1-optical-heart-rate-sensor>

⁹<https://www.maximintegrated.com/en/design/reference-design-center/system-board/6300>



(a) Part of the sensing system: wearable sensors.



(b) IMU sensors integrated into a chair. (c) Heart rate monitor, infrared camera, and EEG headset utilized.

Fig. 2: Experimental testbed.

TABLE I: Sampling rates and missed values for sensor data.

Sensor name	Sampling rate, Hz	Hours collected	Data Missed
EMG sensor			
GSR sensor	36	41.6	0%
IMU right&left hand			
IMU chair seat&back	65	40.1	3.6%
Particle sensor	16-25	41.6	0%
Environmental sensor			
CO ₂ sensor	1	41.6	0%
Eye tracker	90	34.0	18.3%
Mouse&Keyboard	-	35.3	15.1%
Heart rate monitor	1	35.7	14.2%

used three Arduino boards to collect data from each person. EMG and GSR sensors were connected to analog ports on the Arduino 1. Two IMUs on both hands were also connected to Arduino 1 but via I2C protocol. Arduino 2 gathered data from two IMUs located on the chair. Arduino 3 collected the signal from the particle sensor via I2C protocol. Arduino 4 was used to collect environmental data. It received data from the humidity sensor via I2C protocol and from the CO₂ sensor via UART. Arduino boards were powered with standard 5.0V 10000 mAh batteries. In total, we used 16 Arduino boards and 10 batteries.

We used a separate smartphone for each player to manage data collection from a heart rate monitor and an infrared camera. Heart rate monitors communicated with devices via Bluetooth standard; infrared cameras were plugged into Micro-USB or USB-C ports in smartphones. Heart rate monitors were managed by an Android app *Polar Flow*. Infrared cameras were managed by a custom app written using FLIR ONE SDK¹⁰.

Each EEG headset transferred data to a separate PC via Bluetooth standard. Data transmitted were logged into files using a custom python script. Data from each eye tracker were logged to the corresponding gaming PC. Mouse and keyboard activity were also recorded on each gaming PC by a custom python script based on the *pyinput*¹¹ library.

Before each experiment, we synchronized all Arduino boards and PCs with an NTP server. Considering the desynchronization of devices through the experiment's duration, the difference in time did not exceed 500 ms.

In total, we collected 54.6 hours of recordings. Sampling rates for sensors, durations of data collected, and proportions for data missed are presented in Table I. Excluding data from the EEG headset and infrared camera, the duration of data collected is more than 34 hours, and the proportion of data missed does not exceed 18.3%. Duration of the data collected for EEG is 12.2 hours (70.6% missed), for the infrared camera is 26.6 hours (36.1% missed). We present these data in the dataset but don't use it for the analysis. The limitation for EEG data collection was a Bluetooth interference in the room. The limitation for data collection from infrared cameras was a battery life of the devices, because they couldn't be charged when the infrared camera is plugged in.

¹⁰<https://developer.flir.com/mobile/flironesdk/>

¹¹<https://pypi.org/project/pyinput/>

TABLE II: Players' statistics.

Team	Player id	Hours played	Best rank achieved	Current Rank	Rank percentile	Dominant hand
Amateurs	0	336	Gold	Gold	34%	Right
	1	1200	Gold	Gold	34%	Right
	2	1000	Platinum	Platinum	9.66%	Right
	3	400	No rank	No rank	100%	Right
	4	415	Gold	Gold	34%	Right
Professionals	0	5000	Diamond	Diamond	2.07%	Left
	1	6000	Diamond	Diamond	2.07%	Right
	2	5000	Diamond	Diamond	2.07%	Right
	3	10000	Master	Master	0.09%	Right
	4	7000	Diamond	Diamond	2.07%	Right



Fig. 3: Process of data collection.

After each experiment, we aggregated the data collected from Arduino boards, smartphones, and PCs to one storage device.

IV. METHODOLOGY

We invited an amateur team and a professional team to participate in the experiments. When inviting to participate in the experiments, we informed the participants about the data collection process and how their data will be utilized. Each team consists of five players characterised by similar skill level. The information about participants and their experience is shown in Table II. The process of the experiment is shown in Figure 3. Game client versions varied from 9.22.296.5720 to 9.24.300.6382, depending on an experimental day.

Each team played up to 4 matches per day on 3 different days. Each match was played either versus bots or real opponents from the Internet; either with or without the team communication (2x2=4 options in total). On each experimental day, a team played up to 4 modifications in random order. In total, each team played:

- 3 matches vs real opponents with communication.
- 3 matches vs real opponents without communication.
- 3 matches vs bots with communication.
- 2 matches vs bots without communication (not 3 because of a time limitation in day 2).

Match outcomes for both teams are detailed in Tables IIIa and IIIb.

TABLE III: Match outcomes for teams.

	Bots	Real Opponents
With communication	3/3	0/3
Without communication	2/2	1/3

(a) Match outcomes for the amateur team.

	Bots	Real Opponents
With communication	3/3	1/3
Without communication	2/2	2/3

(b) Match outcomes for the professional team.

In bot games, the difficulty of bots was selected to "Intermediate," which was the highest level offered by League of Legends at the time of the experiment. Games versus bots help to establish the baseline to properly compare the amateur and the professional team. Participants won all matches versus bots. In games versus real opponents, a team played against five people from the Internet selected by an in-game matchmaking algorithm. The matchmaking algorithm tries to find opponents similar to participants in terms of skill.

The reason to record games with or without communication is to check the importance of cooperation for teams of different skill. According to the prior research, quality of team communication might significantly impact eSports team performance [45].

After each match had finished, each player passed a short survey and indicated mental load in the game, the level of disturbance and obtrusiveness of the sensing system, and also estimated a personal performance and the performance of teammates.

According to these reports, in 17.5% of matches, players were not concerned about the sensing system at all; in 69.7% of matches, they indicated a little bit of disturbance; in 12.8% of matches, they indicated a significant discomfort caused by the sensing system.

V. DATA COLLECTED

A. Dataset Structure

The dataset consists of data collected in 22 matches and anonymized participants' data. Data about each match are composed of information obtained from in-game logs, sensors, and surveys. The dataset structure is shown in Fig. 4. The dataset consists of 22 directories for each match. Match directories include

- Meta-information file `meta_info.json` specifying a team (amateurs or professionals), type of opponents (bots

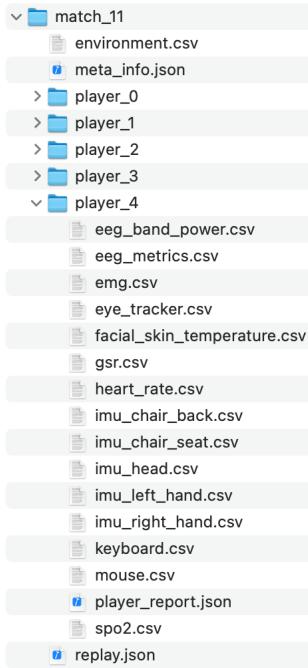


Fig. 4: Dataset structure.

or real players), whether team communication were allowed, a match outcome, match duration, etc.

- Match replay in `replay.json` file with information about game events, such as kill, deaths, items, abilities learned, etc.
- Five directories with sensor data and post-match surveys for each player. Directories are named as `player_ID`, where `ID` is a player id from 0 to 4.
- A file with environmental data `environment.csv`.

Each directory `player_ID` with data collected from a player includes:

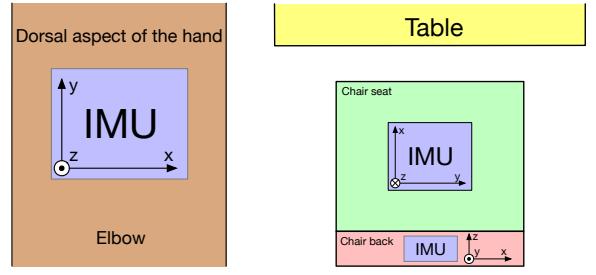
- A number of `.csv` files with sensor data.
- A `player_report.json` file with players' reports on the post-match survey.

B. In-game Data

We obtained in-game logs using the Riot API¹². The data are match timelines, including information about key game events and statistics. Records collected include the following events:

- 1) ***CHAMPION_KILL***. This event indicates the killer, the victim, the assistants, as well as the location and the timestamp.
- 2) ***BUILDING_KILL***. The record specifies the type and the location of the building destroyed, players participated, and the timestamp.
- 3) ***SKILL_LEVEL_UP***. The entry indicates which champion's skill player chooses to upgrade.
- 4) ***WARD_PLACED***, ***WARD_KILL***. These events specify the ward type, the timestamp, and the player involved.

¹²<https://developer.riotgames.com/>



(a) Orientation of IMU on a wrist. (b) Orientation of IMUs on a chair.

Fig. 5: Axes orientation for IMUs located on wrists, chair seat, and chair back.

- 5) ***ITEM_PURCHASED***, ***ITEM_DESTROYED***. The records indicate the item, the timestamp, and the player involved. In-game data are available in the `replay.json` file.

C. Sensor Data

The overview of collected sensor data is shown in Table IV. This table describes which signal is measured and how corresponding files are named in the dataset. In addition to the actual sensor signal, each record includes a relative timestamp of the measurement.

The sensor data collected are presented in Fig. 6. Visualizations of EEG data collected are presented in Appendix A.

These data were processed to remove the noise and outliers. Outliers for each sensor were removed by clipping values to an interval between 0.5 and 99.5 percentiles. Then the signal was smoothed using an exponential moving average with a half-life value of 1 second. Finally, all signals were resampled to a unified timestep of 1 second by averaging or summation, depending on the nature of source sensor data.

D. Self-reported Survey

We collected 109 (out of 110 possible) after-match responses from all players, as described in Section IV. In particular, players were asked to evaluate their own performance as well as the performance of their teammates on a scale from 1 to 5. The distribution of results is shown in Figure 7. Interestingly, players tend to evaluate the performance of their peers higher than their own performance.

Players also reported mental load in the game on a scale from 1 to 5. The average mental load in games versus bots is only 1.2. In matches against real players, the average mental load is about 3.39. The distribution of responses about the mental load in matches against real opponents is shown in Figure 8. Clearly, amateurs' mental load is significantly higher than for professionals, even considering the match-making algorithm paired our participants with similarly skilled players.

VI. DATA ANALYSIS

To show the validity of collected data, we perform the analysis of the dataset. We explore our multi-modal dataset to investigate three eSports research topics regarding the player: skill prediction, player re-identification, and team dynamics.

TABLE IV: Collected Sensor Data.

Sensor Data	Description	Corresponding Files
Hand movements	Data about hand movements were obtained by two IMUs located on both hands and included linear acceleration, gravity, angular velocity, Euler angles, and quaternions. Axes orientation for IMUs is shown in Fig. 5a.	imu_left_hand.csv, imu_right_hand.csv
Chair movements	Chair movements were captured by two IMUs attached to the bottom of the chair seat and the chair back. Data logged include linear acceleration, gravity, angular velocity, Euler angles, and quaternions. Axes orientation for IMUs is shown in Fig. 5b.	imu_chair_seat.csv, imu_chair_back.csv
Head movements	These data are collected by IMU integrated into the EEG headset. Data recorded include linear acceleration, magnetometer data, and quaternions.	imu_head.csv
Gaze	Eyetracker captured the gaze position, pupil diameter, as well as the validity of these data.	eye_tracker.csv
Electrodermal activity	Data about the electrodermal activity are presented in terms of resistance measured in Ohms.	gsr.csv
Muscle activity	Muscle activity is measured by EMG sensor and represented as voltage.	emg.csv
Heart rate	Pulse data were collected by the heart rate monitor on the arm and measured in beats per minute (not the waveform of the pulse oximetry signal).	heart_rate.csv
EEG	The signal is captured by EEG headsets. Data collected include alpha, beta, gamma, and theta waves connected with different types of brain activity. Headsets also provide metrics obtained from EEG data (e.g. Engagement, Stress, Focus).	eeg_band_power.csv, eeg_metrics.csv
Face temperature	Face temperature is presented as the 95-th percentile of values in 48x64 arrays obtained by a thermal camera directed to a player's face. Resulting temperature correspond to player's facial skin temperature.	face_temperature.csv
Keyboard activity	Data about keyboard activity are presented as the number of buttons pressed in the last 5 seconds.	keyboard.csv
Mouse activity	Data about mouse activity are presented as the number of clicks and the distance passed in the last 5 seconds.	mouse.csv
Oxygen saturation	Oxygen saturation is calculated based on reflections of red and infrared light captured by a particle sensor located on the right earlobe.	spo2.csv
Environmental data	Data about temperature, pressure, altitude, humidity, and CO ₂ level. All players share the same recordings of environmental data.	environment.csv

We used python ¹³ package pandas ¹⁴ to preprocess

the data and scikit-learn ¹⁵ for the implementations of

¹³<https://www.python.org/>

¹⁴<https://pandas.pydata.org/>

¹⁵<https://scikit-learn.org/>

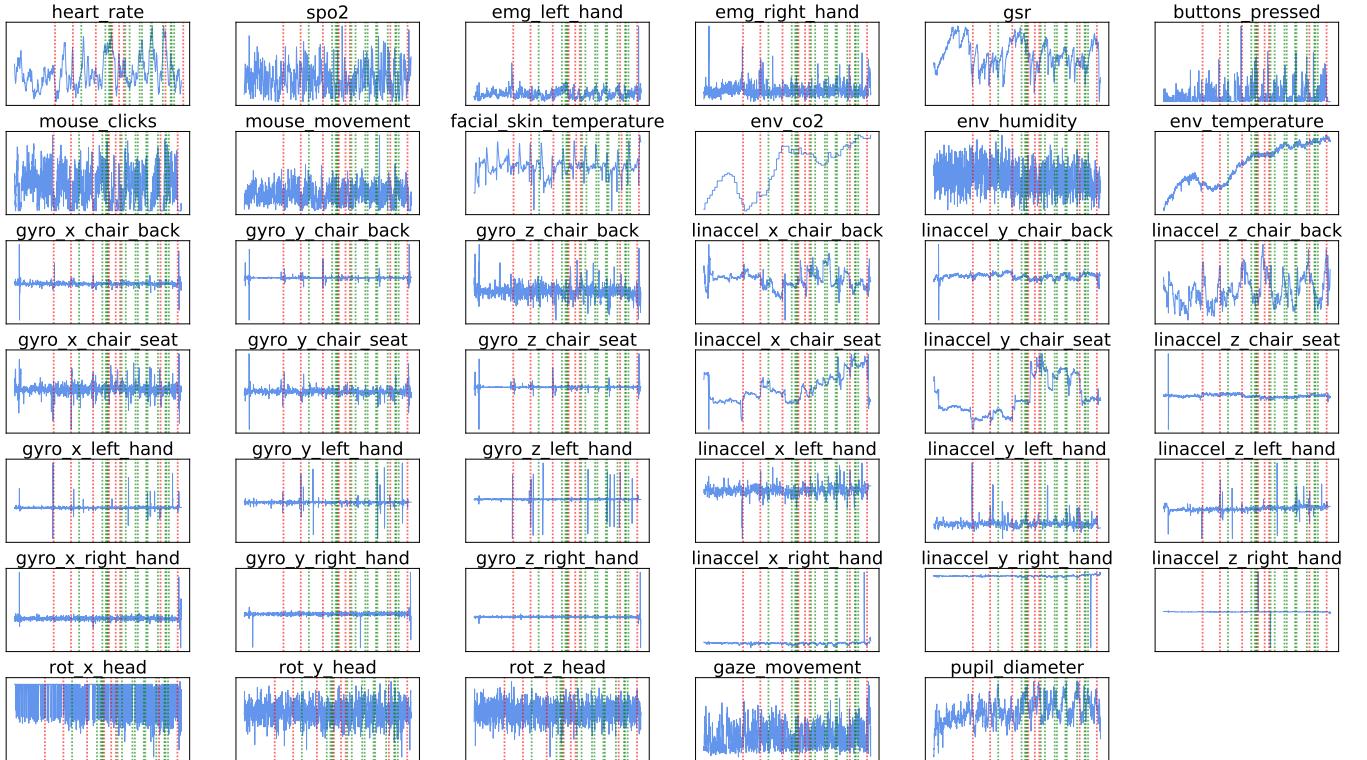


Fig. 6: Selected sensor data collected for one player in one match. Green vertical lines correspond to kill/assist events, red lines correspond to death events.

machine learning models and evaluation metrics.

In the Subsection VI-A we provide results for binary classification of players (pro/amateur) based on sensor data. In the Subsection VI-B we demonstrate the possibility of player-reidentification based on sensor data. Finally, in Subsection VI-C we analyse the differences in team dynamics for the amateur and the professional teams and investigate how lack of communication or playing against real opponents affect it.

A. Skill Prediction

Given the sensor data collected from amateur and professional players, a straightforward question is whether it is possible to distinguish them using only the sensor data. For this purpose, we split data collected for each match into 3-minute sessions, up to 5 sessions per match. In total, we got 540 sessions for both teams. Sessions collected in the first two experimental days were used for training, while sessions from the third experimental day were used for testing.

To extract features from each session, we averaged sensor signal over time and got a 37-dimensional representation for each session (here we didn't use EEG data and environmental data). For each player, each feature in the representation was normalized to zero mean and unit variance. We visualized these multidimensional representations obtained in Fig. 9a and Fig. 9b using 2-dimensional embeddings obtained by the t-SNE method [46]. Fig. 9a is colored w.r.t. to player skill; Fig. 9b is colored w.r.t. to player ID.

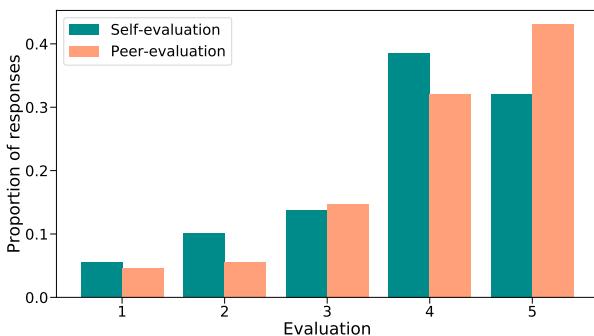


Fig. 7: Distributions of players' responses on teammates performance and self-performance.

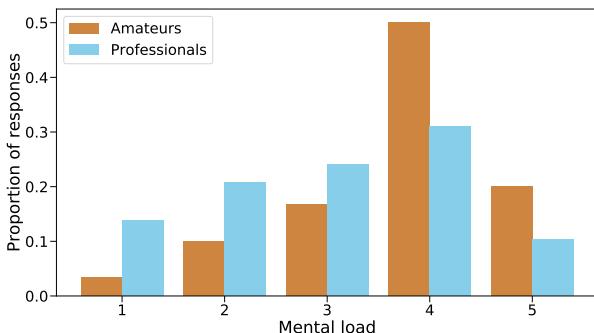
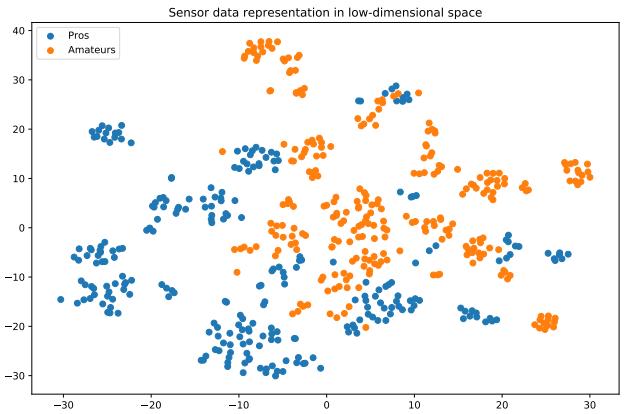
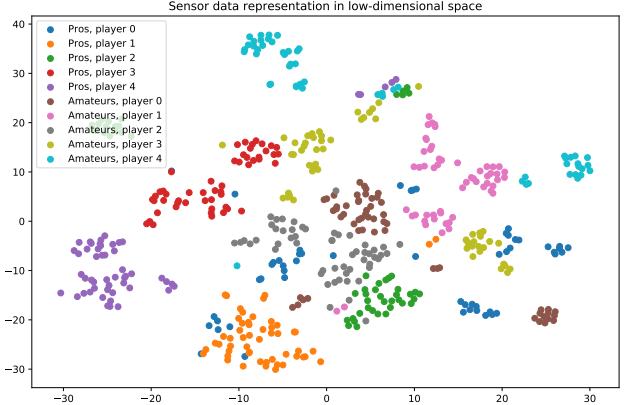


Fig. 8: Distributions of responses on mental load for amateur and professional teams in matches against real opponents.



(a) t-SNE embeddings colored w.r.t. player skill.



(b) t-SNE embeddings colored w.r.t. player ID.

Fig. 9: Sensor data embedded in 2-dimensional space.

Figure 9a demonstrates that sensor data from players with high and low skill levels reside in different regions in the feature space; thus, it may be separated with a classifier.

We trained several classical machine learning algorithms to predict player skill using 37-dimensional feature vectors obtained from sensor data:

- 1) Logistic regression. A simple and robust linear model for classification [47],
- 2) k-nearest neighbors classifier [48] with $k=16$. This algorithm makes predictions based on similarity in feature space. We found $k=16$ to better suit our problem.
- 3) Support Vector Machine with radial basis functions (RBF) kernel. A nonlinear method for robust data classification [49].
- 4) Random forest classifier [50] with 100 estimators and maximum tree depth 4. An ensemble of diverse decision trees.
- 5) Naive Bayes [51]. This method makes predictions assuming that features are independent. We used Gaussian Naive Bayes since sensor data obtained are continuous.
- 6) Gaussian process [52]. This algorithm tries to predict the target class using a latent function.

For evaluation, we used the following metrics:

- 1) Accuracy score [53]. It is calculated as a proportion of right predictions. The higher values are, the better.

TABLE V: Evaluation of machine learning methods for skill prediction. Scores for a random guess are added for comparison.

Method	Accuracy	ROC AUC	Logloss
Logistic regression	0.838	0.886	0.596
k-nearest neighbors	0.741	0.899	0.442
Support vector machine	0.856	0.945	0.311
Random forest	0.800	0.885	0.456
Naive Bayes	0.706	0.780	4.627
Gaussian process	0.791	0.835	0.693
Random guess	0.500	0.500	0.693

- 2) The area under the receiver operating characteristic curve (ROC AUC) [54]. It takes values from 0 to 1 with a 0.5 score for random guessing. The higher values are, the better.
- 3) Log Loss, or binary cross-entropy [55]. This loss function is widely utilized in information theory. The lower values are, the better.

Evaluation results are shown in Table V. We used accuracy, logloss, and ROC AUC metrics for evaluation. The target is balanced, so these metrics are suitable for the problem.

The best evaluation quality is achieved by the SVM method: 0.856 accuracy, 0.945 ROC AUC, and 0.311 logloss scores. In other words, that is possible to estimate player skill by sensor data if the player was presented in a training set. This model can help investigate how player skill changes over time to provide quick feedback and analytics.

B. Player Re-Identification

Another potential application of sensor data is player re-identification. Re-id methods try to identify a player from a database using a data footprint. Such models can be helpful in cheating detection (e.g. when a person plays instead of another player in a tournament), an adaptation of predictive models, or biometric identification. 2-dimensional embeddings obtained by t-SNE (see Fig. 9b) suggest that players may be separated by a classifier.

As our dataset was acquired in a relatively short time span of several weeks, re-id in this case does not reflect the change of playing style or skill improvement over time. But it is interesting to see how players' skill improvement and changes of playstyles can be reflected from the sensor data.

We trained machine learning models on the sensor data described in Section VI-A using player id as a target. Models predicted one of 10 classes corresponding to players. The results are shown in Table VI. We used the One-vs-Rest

TABLE VI: Evaluation of machine learning methods for player re-identification. Scores for a random guess are added for comparison.

Method	Accuracy	ROC AUC	Logloss
Logistic regression	0.488	0.884	1.615
k-nearest neighbors	0.415	0.840	5.735
Support vector machine	0.450	0.890	1.588
Random forest	0.521	0.919	1.617
Naive Bayes	0.341	0.686	22.239
Gaussian process	0.447	0.812	2.302
Random guess	0.100	0.500	2.303

strategy to calculate ROC AUC for multiclass classification [56].

The scores presented show that it is possible to identify a player with 0.521 accuracy, 0.919 ROC AUC, and 1.588 logloss. Relatively high scores for accuracy and ROC AUC demonstrate the possibility of identifying the actual player in the top several predictions.

C. Team Dynamics

A straightforward approach to measure team dynamics in terms of sensor data is to calculate pairwise correlations [57] to estimate how synchronized players physiology. We calculated team dynamics as average pairwise Pearson correlations for different features extracted from sensor data. That means that for each feature we calculated the correlation for each pair of players and then averaged these values for all pairs. The result represents how each feature is synchronized for the whole team. We used `numpy.corrcoef`¹⁶ function to calculate each pairwise correlation. We also estimated errors by checking how results change when calculated on random subsets containing 80% of matches. For all features reported the error doesn't exceed 0.03.

Results are presented in Fig. 10; abbreviations are explained in Table VII.

¹⁶<https://numpy.org/doc/stable/reference/generated/numpy.corrcoef.html>

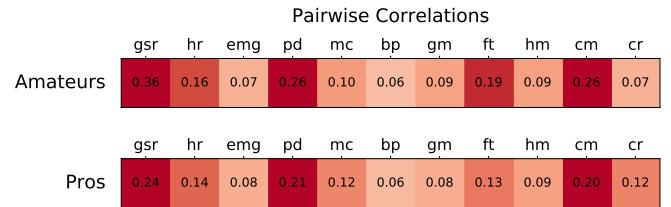


Fig. 10: Team dynamics as average pairwise correlations for sensor data for all matches played.

TABLE VII: Feature Abbreviations.

Abbreviation	Description	Possible Interpretation
gsr	Galvanic skin response	A measure of stress/calmness.
hr	Heart rate	A measure of stress/calmness.
emg	Electromyography	The intensity of the player's muscle activity (here we used data from the right hand).
pd	Pupil diameter	A measure of player's concentration and cognitive load [58], [59].
mp	Mouse clicks	Frequency of mouse clicks.
bp	Buttons pressed	Frequency of keyboard strokes.
gm	Gaze movement	Average saccade speed.
ft	Face temperature	A measure of mental load [60].
hm	Hand linear movement	The intensity of hand movement (here we used data from the right hand).
cm	Chair linear movement	How intensely the player moves in a chair.
cr	Chair rotation	How intensely the player spins on a chair.

In general, physiology for the amateur team is more correlated than for the professional team. That particularly holds for galvanic skin response signal, facial skin temperature, and chair movements. Lower correlations for professionals might imply a more independent playstyle of each player, which is not affected by teammates' mistakes or successes.

All the features are positively-correlated. For both teams, which is self-explanatory, because players are viewing and participating in the same match. However, some features like pupil diameter, galvanic skin response, and chair movements are more correlated than others, and some features like EMG, keyboard activity, and gaze movement are less correlated than others. The possible explanation is that players share a similar level of stress and concentration, although the number and frequency of keystrokes, clicks, and saccades don't correlate much because they are only the medium of controlling the game character.

To investigate the difference between two teams more precisely, we checked how the presence or absence of communication impacts team dynamics. We also compared how games against real people instead of bots change teams' behavior.

Fig. 11a demonstrates the difference between average pairwise correlations in matches with communication and without it. Fig. 11b shows the difference between matches against real opponents and bots.

Communication Added												
	gsr	hr	emg	pd	mc	bp	gm	ft	hm	cm	cr	
Amateurs	+0.17	+0.03	+0.04	-0.01	-0.02	+0.01	+0.02	+0.02	-0.06	-0.10	-0.01	
Pros	-0.01	-0.05	-0.02	-0.02	+0.03	-0.00	+0.01	+0.10	-0.02	+0.08	-0.07	

(a) Changes in average pairwise correlations in matches with allowed communication.

Real Opponents Instead of Bots												
	gsr	hr	emg	pd	mc	bp	gm	ft	hm	cm	cr	
Amateurs	+0.08	+0.05	-0.02	+0.11	+0.01	+0.03	-0.00	+0.12	-0.06	-0.16	+0.04	
Pros	-0.07	-0.05	-0.02	+0.14	-0.00	-0.02	+0.02	+0.01	+0.07	-0.04	+0.02	

(b) Changes in average pairwise correlations in matches with real opponents.

Fig. 11: Changes in team dynamics in different conditions.

When team communication is allowed, galvanic skin response is significantly more correlated for the amateur team, which might indicate that players share similar stress levels. That is not the case for the professional team, which shows more independent play in terms of stress.

In matches against real opponents for both teams, the pupil diameter size is more correlated. Since pupil diameter indicates players' concentration and cognitive load, that implies that players' engagement is more synchronized in matches versus real players rather than versus bots. Facial skin temperature

data for amateurs are also more correlated, while chair movements are less correlated, which is also showing increased concentration in matches against real players.

VII. DISCUSSION

Results in Subsections VI-A, VI-B, and VI-C shows the connection between sensor data and in-game outcomes. These might be utilized in many real-world scenarios.

Skill prediction system might be utilized for recruiting eSports players, evaluating their skill, or providing feedback on which physiological traits can be improved. Player re-identification model might help to detect cheaters. An estimation of team dynamics might help to evaluate how players work as a team. All these data might serve as an additional source of information for the coach.

We hope that both players and community will benefit from these changes, because that will help to bring eSports competition to higher levels.

The results should also hold for other MOBA games (e.g. Dota 2, Heroes of the Storm) because of the similar gameplay, as well as other games of skill (checking this hypothesis is included in our future work). The generalization is achieved by using sensor data instead of data specific to one game. Future work also includes analysing if our findings are applicable in other domains, such as pilot/driver skill estimation, stress monitoring, diseases detection, physical training, etc.

The main limitations of the dataset are:

- Hardware limitations resulting in sometimes incomplete data (e.g. Bluetooth interference for EEG or battery limitations for infrared cameras).
- Only male participants are presented.
- Only 10 players participated in the study.

Future work will focus on these limitations by inviting bigger and diverse set of participants, and using better hardware.

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a dataset with the data from 10 eSports players collected in 22 matches in League of Legends. Data recorded include players' physiology (e.g. hand, head and chair movements, heart rate, gaze, environmental data), in-game events (e.g. kills, abilities, positioning), self-reported surveys from players (prior experience, evaluation of performance), and other metadata. The important aspect of the dataset is simultaneous data collection from 5 players, which enables the consequent analysis of team dynamics. We have compared an amateur and a professional team and learned that professionals stick to a more independent and homogeneous playstyle.

We have also shown the possibility of utilizing sensor data for skill prediction and player re-identification problems. Best models show a validation set accuracy of 0.856 for player skill prediction and 0.521 for player re-id (0.10 for a random guess).

Given the growing popularity of both eSports and wearable devices, utilization of sensor data for video games analytics might provide previously inaccessible feedback and insights,

and in our work we showed that sensor data help understand the difference between the amateurs and professional players. Future work includes a further in-depth analysis of the dataset and investigation of current and new research directions. The multi-modal data collected might be used for many applications, such as encounter outcome prediction, anomaly detection, and performance prediction. We encourage the research community to propose new problems and explore insights based on the dataset collected.

ACKNOWLEDGMENT

The reported study was partially funded by RFBR (Russian Foundation for Basic Research) according to the research project No. 18-29-22077.

The research reported in this paper has been partially supported by the BMBF (German Federal Ministry of Education and Research) in the project HeadSense (project number 01IW18001). The support is gratefully acknowledged.

Also, the authors would like to thank the eSports club¹⁷ at the University of Kaiserslautern for their contribution to methodology development and participation in data collection.

REFERENCES

- [1] V. J. Hodge, S. M. Devlin, N. J. Sephton, F. O. Block, P. I. Cowling, and A. Drachen, "Win prediction in multi-player esports: Live professional match prediction," *IEEE Transactions on Games*, November 2019. [Online]. Available: <http://eprints.whiterose.ac.uk/152931/>
- [2] N. Kinkade, L. Jolla, and K. Lim, "Dota 2 win prediction," *Univ Calif*, vol. 1, pp. 1–13, 2015.
- [3] A. L. C. Silva, G. L. Pappa, and L. Chaimowicz, "Continuous outcome prediction of league of legends competitive matches using recurrent neural networks," in *SBC-Proceedings of SBCGames*, 2018, pp. 2179–2259.
- [4] G. Choi and M. Kim, "Eye gaze information and game level design according to fps gameplay beats," *Journal of information and communication convergence engineering*, vol. 16, pp. 189–196, 2018.
- [5] N. Khromov, A. Korotin, A. Lange, A. Stepanov, E. Burnaev, and A. Somov, "Esports athletes and players: a comparative study," *IEEE Pervasive Computing*, vol. 18, no. 3, pp. 31–39, 2019.
- [6] P. M. Blom, S. Bakkes, and P. Spronck, "Towards multi-modal stress response modelling in competitive league of legends," in *2019 IEEE Conference on Games (CoG)*, 2019, pp. 1–4.
- [7] D. Koposov, M. Semenova, A. Somov, A. Lange, A. Stepanov, and E. Burnaev, "Analysis of the reaction time of esports players through the gaze tracking and personality trait," in *2020 IEEE 29th International Symposium on Industrial Electronics (ISIE)*, 2020, pp. 1560–1565.
- [8] A. Smerdov, E. Burnaev, and A. Somov, "esports pro-players behavior during the game events: Statistical analysis of data obtained using the smart chair," in *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2019, pp. 1768–1775.
- [9] Z. Minchev, G. Dukov, and S. Georgiev, "Eeg spectral analysis in serious gaming: An ad hoc experimental application," *Bio Automation*, vol. 13, no. 4, pp. 79–88, 2009.
- [10] R. Kaminsky, M. Enev, and E. Andersen, "Identifying game players with mouse biometrics," *University of Washington. Technical Report*, 2008.
- [11] B. B. Velichkovsky, N. Khromov, A. Korotin, E. Burnaev, and A. Somov, "Visual fixations duration as an indicator of skill level in esports," in *IFIP Conference on Human-Computer Interaction*. Springer, 2019, pp. 397–405.
- [12] G. Choi and M. Kim, "Eye-movement pattern by playing experience in combat system of fps game," pp. 52–56, 2015.
- [13] B. Meneses-Claudio and A. Roman-Gonzalez, "Study of the brain waves for the differentiation of gamers category between a newbie and a hardcore in the game dota 1," in *2018 Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI)*, 2018, pp. 1–4.
- [14] A. Smerdov, A. Kiskun, R. Shaniazov, A. Somov, and E. Burnaev, "Understanding cyber athletes behaviour through a smart chair: Cs:go and monolith team scenario," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019, pp. 973–978.
- [15] C. Ringer, J. A. Walker, and M. A. Nicolaou, "Multimodal joint emotion and game context recognition in league of legends livestreams," in *2019 IEEE Conference on Games (CoG)*, 2019, pp. 1–8.
- [16] V. Hodge, S. Devlin, N. Sephton, F. Block, A. Drachen, and P. Cowling, "Win prediction in esports: Mixed-rank match prediction in multi-player online battle arena games," *arXiv preprint arXiv:1711.06498*, 2017.
- [17] Y. Yang, T. Qin, and Y.-H. Lei, "Real-time esports match result prediction," *arXiv preprint arXiv:1701.03162*, 2016.
- [18] L. Lin, "League of legends match outcome prediction."
- [19] K. J. Shim, K.-W. Hsu, S. Damania, C. DeLong, and J. Srivastava, "An exploratory study of player and team performance in multiplayer first-person-shooter games," in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 2011, pp. 617–620.
- [20] L. C. Kho, M. S. M. Kasihmuddin, M. Mansor, S. Sathasivam *et al.*, "Logic mining in league of legends." *Pertanika Journal of Science & Technology*, vol. 28, no. 1, 2020.
- [21] Z. Yang, Z. Pan, Y. Wang, D. Cai, S. Shi, S.-L. Huang, and X. Liu, "Interpretable real-time win prediction for honor of kings, a popular mobile moba esport," *arXiv preprint arXiv:2008.06313*, 2020.
- [22] M. Schubert, A. Drachen, and T. Mahlmann, "Esports analytics through encounter detection," in *Proceedings of the MIT Sloan Sports Analytics Conference*, vol. 1. MIT Sloan Boston, MA, 2016, p. 2016.
- [23] D.-H. Kim, C. Lee, and K.-S. Chung, "A confidence-calibrated moba game winner predictor," *arXiv preprint arXiv:2006.15521*, 2020.
- [24] A. Semenov, P. Romov, S. Korolev, D. Yashkov, and K. Neklyudov, "Performance of machine learning algorithms in predicting game outcome from drafts in dota 2," in *International Conference on Analysis of Images, Social Networks and Texts*. Springer, 2016, pp. 26–37.
- [25] N. Wang, L. Li, L. Xiao, G. Yang, and Y. Zhou, "Outcome prediction of dota2 using machine learning methods," in *Proceedings of 2018 International Conference on Mathematics and Artificial Intelligence*, 2018, pp. 61–67.
- [26] Z. Chen, T.-H. D. Nguyen, Y. Xu, C. Amato, S. Cooper, Y. Sun, and M. S. El-Nasr, "The art of drafting: a team-oriented hero recommendation system for multiplayer online battle arena games," in *Proceedings of the 12th ACM Conference on Recommender Systems*, 2018, pp. 200–208.
- [27] A. White and D. M. Romano, "Scalable psychological momentum forecasting in esports," *arXiv preprint arXiv:2001.11274*, 2020.
- [28] Y. Yang, T. Qin, and Y.-H. Lei, "Real-time esports match result prediction," *arXiv preprint arXiv:1701.03162*, 2016.
- [29] I. Makarov, D. Savostyanov, B. Litvyakov, and D. I. Ignatov, "Predicting winning team and probabilistic ratings in ‘dota 2’ and ‘counter-strike: Global offensive’ video games," in *International Conference on Analysis of Images, Social Networks and Texts*. Springer, 2017, pp. 183–196.
- [30] A. Dehpanah, M. F. Ghori, J. Gemmell, and B. Mobasher, "The evaluation of rating systems in online free-for-all games," *arXiv preprint arXiv:2008.06787*, 2020.
- [31] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau, "Guns, swords and data: Clustering of player behavior in computer games in the wild," in *2012 IEEE conference on Computational Intelligence and Games (CIG)*. IEEE, 2012, pp. 163–170.
- [32] L. Gao, J. Judd, D. Wong, and J. Lowder, "Classifying dota 2 hero characters based on play style and performance," *Univ. of Utah Course on ML*, 2013.
- [33] J. Pfau, J. D. Smeddinck, I. Bikas, and R. Malaka, "Bot or not? user perceptions of player substitution with deep player behavior models," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–10.
- [34] S. Adinolfi and S. Turkay, "Toxic behaviors in esports games: player perceptions and coping strategies," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 2018, pp. 365–372.
- [35] M. Martens, S. Shen, A. Iosup, and F. Kuipers, "Toxicity detection in multiplayer online games," in *2015 International Workshop on Network and Systems Support for Games (NetGames)*, 2015, pp. 1–6.

¹⁷<https://www.esports-kl.de/>

- [36] P. M. Lehrer, D. M. Batey, R. L. Woolfolk, A. Remde, and T. Garlick, "The effect of repeated tense-release sequences on emg and self-report of muscle tension: An evaluation of jacobsonian and post-jacobsonian assumptions about progressive relaxation," *Psychophysiology*, vol. 25, no. 5, pp. 562–569, 1988.
- [37] D. C. Fowles, "The three arousal model: Implications of gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy," *Psychophysiology*, vol. 17, no. 2, pp. 87–104, 1980.
- [38] M. Zhu, W. Liu, and P. Wargoocki, "Changes in eeg signals during the cognitive activity at varying air temperature and relative humidity," *Journal of Exposure Science & Environmental Epidemiology*, vol. 30, no. 2, pp. 285–298, 2020.
- [39] L. Lan, P. Wargoocki, D. P. Wyon, and Z. Lian, "Effects of thermal discomfort in an office on perceived air quality, sbs symptoms, physiological responses, and human performance," *Indoor air*, vol. 21, no. 5, pp. 376–390, 2011.
- [40] J. G. Allen, P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, and J. D. Spengler, "Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments," *Environmental health perspectives*, vol. 124, no. 6, pp. 805–812, 2016.
- [41] G. Stenberg, "Personality and the eeg: Arousal and emotional arousability," *Personality and individual differences*, vol. 13, no. 10, pp. 1097–1113, 1992.
- [42] D. Buckley, K. Chen, and J. Knowles, "Predicting skill from gameplay input to a first-person shooter," in *2013 IEEE Conference on Computational Intelligence in Games (CIG)*. IEEE, 2013, pp. 1–8.
- [43] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," in *4th European conference of the international federation for medical and biological engineering*. Springer, 2009, pp. 1366–1369.
- [44] A. B. Scholey, M. C. Moss, and K. Wesnes, "Oxygen and cognitive performance: the temporal relationship between hyperoxia and enhanced memory," *Psychopharmacology*, vol. 140, no. 1, pp. 123–126, 1998.
- [45] M. J. Smith, P. D. Birch, and D. Bright, "Identifying stressors and coping strategies of elite esports competitors," *International Journal of Gaming and Computer-Mediated Simulations (IJGEMS)*, vol. 11, no. 2, pp. 22–39, 2019.
- [46] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [47] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.
- [48] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers—," *arXiv preprint arXiv:2004.04523*, 2020.
- [49] S.-i. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.
- [50] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [51] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.
- [52] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Springer, 2003, pp. 63–71.
- [53] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," in *Australasian joint conference on artificial intelligence*. Springer, 2006, pp. 1015–1021.
- [54] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (roc) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19–20, 2006.
- [55] V. Vovk, "The fundamental nature of the log loss function," in *Fields of Logic and Computation II*. Springer, 2015, pp. 307–318.
- [56] J. Hanley and B. McNeal, "A simple generalization of the area under the roc curve to multiple class classification problems," *Radiology*, vol. 143, pp. 29–36, 1982.
- [57] E. R. Walker, D. L. Lang, B. A. Caruso, and L. Salas-Hernández, "Role of team dynamics in the learning process: a mixed-methods evaluation of a modified team-based learning approach in a behavioral research methods course," *Advances in Health Sciences Education*, pp. 1–17, 2019.
- [58] O. E. Kang, K. E. Huffer, and T. P. Wheatley, "Pupil dilation dynamics track attention to high-level information," *PloS one*, vol. 9, no. 8, p. e102463, 2014.
- [59] B. Stone, M. Lee, S. Dennis, and T. Nettelbeck, "Pupil size and mental load," in *1st Adelaide Mental Life Conference*, 2004.
- [60] T. Mizuno, T. Sakai, S. Kawazura, H. Asano, K. Akehi, S. Matsuno, K. Mito, Y. Kume, and N. Itakura, "Measuring facial skin temperature changes caused by mental work-load with infrared thermography," *IEEJ Transactions on Electronics, Information and Systems*, vol. 136, no. 11, pp. 1581–1585, 2016.

APPENDIX A EEG DATA

Additional visualizations for EEG band power and EEG performance metrics are shown in Fig. 12 and Fig. 13.

EEG band power represents the intensity of include alpha, beta, gamma, and theta waves connected with different types of brain activity.

EEG performance metrics are calculated by the headset¹⁸ and provides six basic measures of mental performance. Each measure is automatically scaled to suit participant's normal range and base level of each condition.

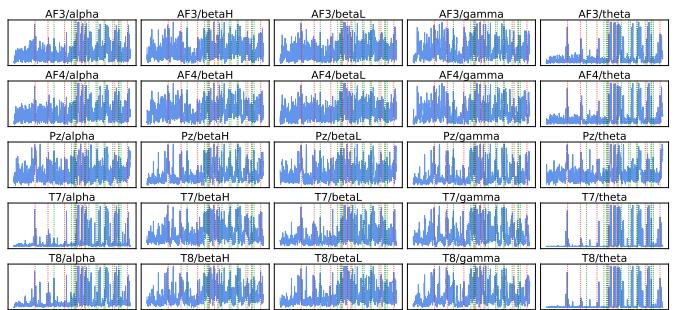


Fig. 12: EEG band power for one match for one player. Green vertical lines correspond to kill/assist events, red lines correspond to death events.

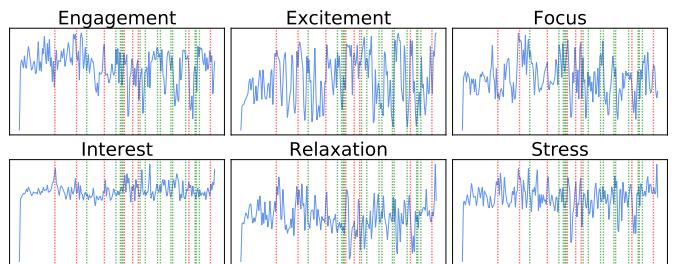


Fig. 13: EEG metrics calculated by the EEG headset for one match for one player. Green vertical lines correspond to kill/assist events, red lines correspond to death events.

¹⁸<https://www.emotiv.com/knowledge-base/performance-metrics/>