Classification of Physiological Data in Affective Exergames

Andreas Kamenz, Victoria Bibaeva
HAW Hamburg, Informatik
Hamburg, Germany
andreas.kamenz@haw-hamburg.de

Arne Bernin
HAW Hamburg, Informatik
University of the West of Scotland (UWS)
Hamburg, Germany

Sobin Ghose
HAW Hamburg, Informatik
Hamburg, Germany
sobin.ghose@haw-hamburg.de

Kai von Luck
HAW Hamburg, Informatik
Hamburg, Germany
luck@informatik.haw-hamburg.de

Florian Vogt
Innovations Kontakt Stelle (IKS)
HAW Hamburg, Informatik
Hamburg, Germany

Larissa Müller
HAW Hamburg, Informatik
PricewaterhouseCoopers GmbH WPG
Hamburg, Germany
larissa.k.mueller@googlemail.com

***Abstract*—In this work, we present our approach to analyze physiological data in affective exergames with deep learning algorithms. In previous works, a cycling exercise machine has been enhanced to act as a game controller. During a case study, we collected vision-based and physiological data of 25 participants who rode through the designed emotion provoking game environment. Using the collected physiological data, we propose to use ensemble learning based on three distinct deep learning models: Multilayer Perceptron, Fully Convolutional Networks and Residual Networks. The proposed algorithms were able to enhance our previously introduced event-based emotion analysis method.**

***Index Terms*—Affective Computing, Exergames, Physiological Data, Deep Learning, Time Series Classification**

I. Introduction

The influence of affective computing on various industries has grown during the last decades. Nowadays, marketing strategies often include an evaluation of customers' emotions, and the automotive industry analyzes users' attention. Modern games aspire to integrate user emotions in their game play [9]. Since emotions are able to increase motivation [1], the sports industry can also highly benefit from affective computing technologies. Moreover, it has been shown that entertaining content can provide motivation for physical exercise [2].

Designing entertainment systems is very challenging, since the perception of entertaining values and the emotional reaction to similar game elements is highly individual. One consequence is that a different personal experience is created for each user. Therefore, the presented system is designed to analyse these individual reactions by modern deep learning approaches.

Nowadays, many affective algorithms are well tested and approved in desktop scenarios. Less research is devoted to interactive exercising context, since such a scenario requires movement and physical effort. The analysis is more complex, and therefore movement is seldom included in the acquisition [7]. We find it interesting since it goes beyond classical desktop approaches and it has been shown that emotions are linked with physical activity [11], related to happiness [10] and that they are able to enhance exercise experiences [12].

In previous works, a physical cycling exergame was presented and facial expressions were analyzed in an exergaming context [3]. In another work, facial expressions were combined with physiological data to enhance emotion recognition rates [4]. In further works, we enhanced the system by a dynamic game play based on the emotional reactions in order to steer participants on a predefined path of emotions [6]. Thereby a database was created which contains of video and thermal camera data as well as physiological time series data consisting of EDA, respiration and temperature sensors. EDA is a part of the autonomous nervous system and thereby known to be closely associated with the arousal of participants [15] and allows to recognize basic emotion as well [16]. In this work, we focus on improving the emotion recognition system by a deeper analysis of the physiological data.

II. Related Work

As opposed to other classification problems, Time Series Classification (**TSC**) deals with data which is ordered by time. This contraint imposes that the discriminative features extracted from this kind of data also depends on ordering [22].

The solutions to TSC problems can be grouped into shape-based, feature-based and direct approaches [23].

Shape-based approaches attempt to measure the similarty of each pair of time series samples based on their shape, which allows them to use the similarity to classify new time series samples. The most prominent and commonly used algorithm of this group is Dynamic Time Warping (DTW) combined with a classifier such as k-Nearest-Neighbours [22].

Feature-based approaches extract a set of statistical features from the time series data like maximum, minimum, skewness, curtosis, peaks etc. Many of these features are designed specifically for certain kinds of data, e.g. logarithmic periodogram for contact strips in trains [23]. The researchers then have to define a wide range of generic features to be calculated, before the actual classification takes place.

As an alternative to hand-crafted features, feature learning is the basis of direct approaches to TSC problem. This group of approaches must take the above mentioned locality of features into account. Direct algorithms are especially suitable for problem instances where manual feature-engineering is very hard [23]. Recently proposed instances of direct algorithms are recurrent and convolutional neural networks as well as stacked restricted Boltzmann-machines (for references see [23] and [19]). They belong to the deep learning domain.

Deep learning is becoming more and more popular as it achieves strongly competitive results in many signal processing tasks such as image and audio classification, video analysis etc. Due to their increasing prevalence, deep learning models are successfully transferred to other research areas, including TSC. Thus, we investigate the use of deep learning models in an ensemble for the practical TSC task at hand.

Our problem setting is complicated by the fact that the given data is spread across different channels, each of which is produced by one bio-sensor. It implies using some kind of data fusion. The authors of [18] faced similar problem with different data channels (namely video, audio and text) for the deception detection task. The solution was to test two different fusion mechanisms. The first mechanism, *feature-level fusion*, concatenates the extracted features before feeding them to a classifier. The second, *decision-level fusion*, employs several so-called "weak learners" that learn features from each data channel separately. Afterwards, the class predictions are calculated out of all weak learners' outputs, which corresponds to the general ensemble learning procedure. Similar comparison of data fusion mechanisms was done in [17], where the goal was to classify physiological time series data for emotion recognition. The results have shown the superiority of the decision-level fusion, which we have therefore chosen for our experiments.

Our work differs from the mentioned references in that we explore the emotion recognition task in context of exergames, which implies collecting and analyzing physiological data. Our assumption is that emotion recognition based in physiological data can be achieved through creating an exergame environment where the participants are actively influence the virtual bicycle, at the same time facing different objectives and being emotionally involved.

Whereas the authors of [17] utilize the already available MAHNOB-HCI database, where 27 participants are sitting while being shown a set of emotion provoking videos, our trial setting involves the participants' movement. Also, the study [17] uses other kinds of weak learners, namely AdaBoosted Trees Classifiers, whose outputs are combined in order to derive the final results. In contrast, we propose ensemble learning for sensor fusion, where the weak learners are deep neural networks. These networks have already shown excellent results on other TSC datasets [19], hovewer being used as single models for time series containing a single channel.

III. System Setup

The system setup is designed as a testbed that supports the evaluation of emotion recognition and emotion provocation technologies in an exercising context. Furthermore it enables to conduct experiments easily to test miscellaneous methods. It contains a physical exergame controller, a data acquisition system, different emotion sensors and an emotion provoking game. Furthermore the participants was asked to fullfil a couple of experimental trials and a database was created, that contains all the sensor data, a screen dump of the virtual game as well as the data logs of the physical cycling exergame controller.

A. Physical Exergame Controller

The exergame system setup includes a cycling exercise machine, modified by a rotatable handlebar, a gear shift and a brake to act as a physical game controller for a specially designed emotion provoking cycling game [5]. The user's revolutions per minute (rpm) while exercising was transmitted to the system to calculate the speed of the virtual bicycle. Thereby, the game is controlled by the player, who has to physically accelerate and steer through the designed virtual environment, that was presented to the user on a display in front of the bike. The designed virtual bicycle had soft real time requirements since insufficient controls may influence the user perception and provoke frustration.

B. Data Acquisition System

A data acquisition system was designed to collect data from different emotion sensors [6]. It is designed as a distributed system with loosely coupled client computers, connected with a message broker (Apache ActiveMQ[1]) and a JSON-based protocol.

A Microsoft Kinect v2 camera[2] was placed in front of the exercise machine to collect video data of the participant's face. It provides HD(1080P) resolution and RGB-D(512x424) images at up to 30Hz. The physiological data was applied by the wearable sensing platform biosignalsplux[3]. The platform provides EDA, blood volume pulse (BVP), ECG, piezoelectric respiration, and temperature sensors and operates at a sampling rate of 256Hz. All the sensors are connected to the Plux hub, sending the data via Bluetooth to our server.

C. Exergame Design

The presented exergame was designed in accordance to the requirements for affective games described in related research [8]: they should be engaging, intuitive, easy to learn, highly dynamic and enable multiple forms of adaptation. Different game scenes enables multiple forms of adaptation and each scene was tailored to steer the participant in a controlled emotional state. The provocation of the different game scenes was evaluated in a previous work [5]. In this work we focus on a subset of the emotion provoking game events that have been

[1] http://activemq.apache.org/

[2] https://developer.microsoft.com/en-us/windows/kinect

[3] http://biosignalsplux.com/

shown to provoke strong physiological responses and thereby are particularly suitable for deep learning approaches. In this work we focus on the analysis of the scenes and events as presented in Table I. A detailed scene description can be found in [5] an [6]. In this work we present a short introduction to the relevant scenes.

1) Challenging Scenes: In the *Challenge Scene* the participants have to fulfill a very challenging task. They need to steer through a booster gate to achieve the right speed and land on the other side of a mountain gap, as shown in Figure 1.

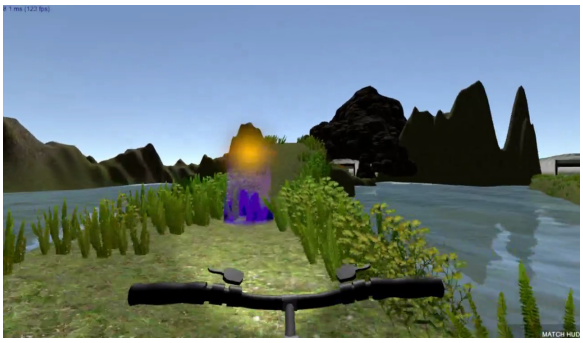

Fig. 1. Booster in the *Challenge Scene*

In the *Downhill Scene* the participants ride down a hill on a wooden path. The speed of the virtual bike is increased due to the negative slope. In front of the finish line a sharp curve needs to crossed in spite of the high speed, which makes this task very challenging.

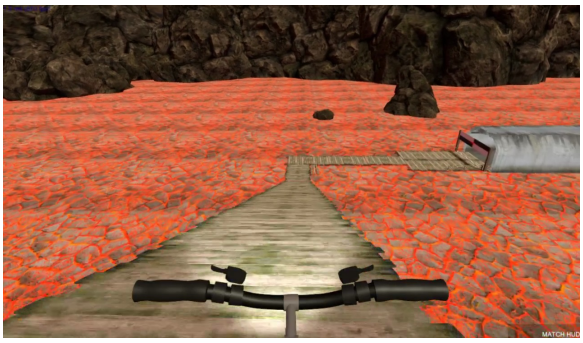

Fig. 2. Challenging Task in the *Downhill Scene*

2) Scary Scenes: In this work we focus on the *Forest Scene* and the *Explosion Scene*. In both scenes an uncomfortable environment needs to be crossed. In the *Forest Scene* the player has to take a ride through a dark forest and on its way a *Jump Scare Event* with shocking Zombies. In the *Explosion Scene* a war zone is presented to the participants, as shown in Figure 3. While riding through the scene surprisingly explosions occurs with a loud sound.

D. Exergame Trials

Each trial started with a training scene to allow the user to become familiar with the controls. After that, a dynamic sequence of game scenes was started as described in [5] and

Fig. 3. War Zone in the *Explosion Scene*

TABLE I
OVERVIEW OF EMOTIONAL PROVOCATION IN GAME SCENES

Scene	Event	Objective	Target Emotion	Target Reaction
Challenge	Falling	Make it across the finish line	Joy, Frustration	EDA response
Downhill	Falling	Make it across the finish line	Joy, Frustration	EDA response
Forest	Jump Scare	Cross a dark forest	Surprise, Fear	EDA response
Explosion	Explosion	Cross a war zone	Surprise, Fear	EDA response

took the user on an *Emotional Journey* [6]. The trials were designed to provide a moderate strain to the participants. The personal perception of the physical strain was evaluated by the Borg scale ranging from 1 to 7 [13].

The sensor data acquisition should be minimally inversive to the participants and to avoid feelings of discomfort the sensor mounting was supported by an experimenter of the same gender as the participant. Otherwise they might influenced by undesired emotions. One of our aims was to collect a database that was created during a case study with 25 participants, ten males and fifteen females. Since the annotation of emotions is still an ongoing challenge and it is not certain whether a self-assessment is sufficient or if an observer-assessment is preferable [14] we decided to validate our results by a self-assessment and an observer-assessment.

E. Dataset

In order to achieve our goal to classify physiological data with respect to human emotions we manually created the time series dataset. This was done after the trials with the participants by manually selecting the representative time series and the corresponding emotion labels.

Since all the scenes and emotion provoking events were stored in a database with a timestamp, the challenge was to assign an emotion label to a certain time window that surrounded the event timestamp and contained the relevant sensor data. We could build up on the previously introduced event-based analysis method [4], defining a window of 1 second before each event occurs and 10 seconds after the

event. Our findings show that the selected time window is sufficient to analyze the physiological reactions on every event.

Additionally, we chose the game scenes with the strongest physiological reactions – see Table I. Thus, the resulting dataset includes 382 time series that belong to 2 separate emotional classes. Each time series contains of 5 separate channels, one channel per sensor, downsampled to 128 Hz. One time series includes 1408 sensor values per channel.

Due to the noticeable variations in psycho-physiological qualities of the participants, we normalized each time series channel-wise, subtracting the minimal value and dividing by the value range, as in [17].

IV. Our Approach

We propose to use three deep learning models trained in ensemble to classify the physiological time series data. These are popular models from the deep learning domain, which have already been used for time series classification with many benchmark datasets as described in [19]. The first model is multilayer perceptron (**MLP**), the second is fully convolutional network (**FCN**), the third is residual network (**ResNet**).

The advantages of the proposed models are not only their ability to utilize the input data without much preprocessing, but also feature learning instead of manual feature extraction. Also, according to study [19], these models were able to beat many of shape- and feature-based methods on a large collection of time series benchmark datasets. Finally, due to their depth the proposed models are capable of coping with the high input data complexity [21] that can be expected of physiological data.

The novelty of our approach is to use an ensemble in order to eliminate the uncertainty of choosing, which channel/sensor is the most reliable source to predict emotions. Furthermore, ensemble learning allows us to use all the available sensor information under the assumption that the latter produces higher recognition rates than any of the sensors alone.

A. Multilayer Perceptron

MLP is the most common type of machine learning models that consists of multiple layers of artificial neurons [20]. In our experiments, we chose to use an MLP with 3 fully-connected hidden layers, each having 500 neurons, followed by a softmax layer. Typical techniques such as rectifying non-linearity and dropout were used as well, the former speeding up the training, the latter preventing MLP from overfitting (for definitions see [21]). The probability of dropout before the input layer was set to 0.1, before each hidden layer – to 0.2, and before the softmax layer – to 0.3 [19].

B. Fully Convolutional Network

Our next model, FCN, is a special type of convolutional neural networks (**CNNs**). In turn, CNN is a variant of MLP with some distinctive types of neuron layers, such as convolution layer, pooling layer and full-connection layer (cf. [20]). These layer types are stacked together to form a linear network architecture that efficiently learns feature hierarchies with help of much less neuron weights than MLP. CNN is typically used for data which exhibits some form of local dependencies [21].

In contrast to CNN, FCN does not use any pooling and full-connected layers, instead making advantage of batch normalization after each convolution layer as well as of global pooling at the last layer. Main benefits of these choices are the reduced network size and faster convergence (see [19]).

Thus, our FCN model is composed out of 3 stacked modules, which in turn contain convolution layer, batch normalization layer and rectifying non-linearity. The convolution filters are initialized with the Glorot Scheme [24], filter sizes being 8, 5 and 3, respectively. Filter count of the first and last modules was fixed to 128, as a factor of the time series length, the second module having 256 filters. Again, the last layers of the FCN are global pooling layer and softmax layer, which is responsilble for class prediction.

C. Residual Network

The last model, ResNet, was designed by [25] to overcome the problems of training very deep neuronal networks, e.g. with 152 layers. The solution was to introduce shortcut connections that perform identity mapping without added complexity, resulting in deeper models with better accuracy. Our ResNet is built out of 3 identical residual blocks. Every block has 3 convolution layers with filter sizes 8, 5 and 3, all of which are followed by batch normalization and rectifying non-linearity. The shortcut connection is stretched from the block input to its output. Beside this fact, the only difference between a residual block and FCN described above is the filter count, in the current case being 64, 128 and 128, respectively, to allow stacking of blocks with enough complexity. Finally, the last layers of ResNet are the same as in FCN.

D. Network Ensembles

As stated above, the proposed deep learning models can be used to classify one data channel at a time. In order to combine the class predictions coming from each data channel, an ensemble approach is now introduced.

We designed and compared 3 different ensembles in this paper. Each ensemble contains of 5 equal models working with its own data channel. The first ensemble contains of MLPs, the second – of FCNs, the third – of ResNets. The entire ensemble architecture is illustrated in Figure 4 The single models are represented with FS_i blocks, their outputs h_i being averaged to produce an ensemble output.

V. Experimental Results

We implemented our models in Python using Keras[4] as a training framework. The dataset was split into training and test set with a ratio of 70% to 30%. All models were trained with AdaDelta algorithm. The initial training parameters like learning rate are described in [19]. Additionally, we used random seeds to achieve deterministic results for each run. We conducted 10 runs in total with different random seeds.

[4]https://keras.io/

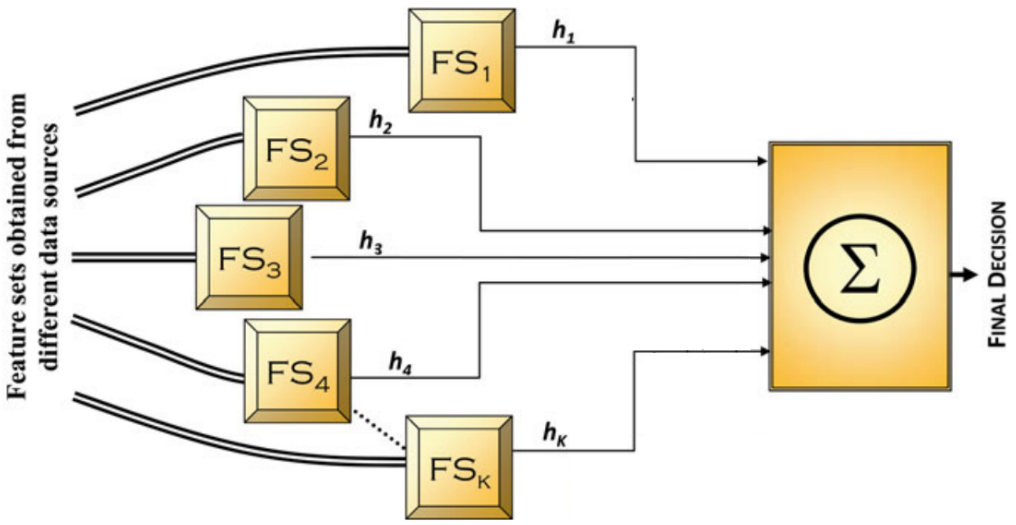

Fig. 4. Deep model ensemble used in this paper (derived from [26]).

The overall results including mean accuracy and standard deviation over 10 runs, are illustrated in Figure 5. The average accuracy of the MLP ensemble is 0.6273, meaning that 62.73% of time series are correctly classified. The average accuracy of FCN ensemble is 0.6559, and of ResNet ensemble – 0.6534. Furthermore, Figure 5 contains the accuracies of each model. In most cases, the highest accuracy was achieved by FCN, including the case of ensemble model. Unsurprisingly, MLP showed the worst accuracy as it is prone to overfitting.

Consequently, we are able to compare the accuracy of each weak learner with the overall ensemble accuracy. This results prove our assumption that the accuracy of an ensemble is higher than the accuracy of each single model.

The analysis of BVP shows the highest rates. This is an interesting finding, since the previous analysis method was based on EDA data. The evaluation of BVP data is promising to enhance the successful recognition rates and thereby should be integrated into the exergame system. Figure 6 shows a typical time series of BVP sensor data in the *Downhill Scene*.

VI. Conclusion

This paper deals with the problem of emotion recognition based on physiological data in context of exergames. In previous research, we collected sensor data under the assumption that they can be used to classify emotional states. We approached the emotion classification as a time series classification problem and proposed three solutions based on well-known deep learning models: Multilayer Perceptron, Fully Convolutional Network and Residual Network. These models were used in an ensemble, which allowed us to use all sensor data separately and later combine them to improve the prediction rate.

Our experiments illustrate the successful use of deep learning models for such non-trivial tasks as emotion recognition in exergames. The proposed models were able to achieve a decent accuracy level despite high data complexity and turned our attention to certain sensors that produce machine distinguishable data. Moreover, we compared the proposed ensembles against each other as well as against single channel models. The most promising results were obtained with FCN ensemble, leading to the necessity of further research to increase its performance. The lowest accuracy and a tendency for overfitting was seen in MLP ensemble. Finally, each ensemble was capable to beat the single channel models, stressing the necessity to employ ensemble learning for such tasks.

Our future research focus will be to investigate the influence of preprocessing techniques, which may lead to an increasing accuracy. Also, better results may be achieved through the use of similar benchmark datasets containing physiological data such as MAHNOB-HCI. These datasets can be used to train the proposed ensembles, after which the models can be fine-tuned with our own collected dataset. Concluding our paper, we improved our previous results on emotion recognition and will continue to research this topic even further.

References

[1] Veronika Brandstätter, Julia Schüler, Rosa Maria Puca, and Ljubica Lozo, "Motivation und Emotion: Allgemeine Psychologie für Bachelor," Springer, 2013.

[2] Rainer Malaka, "How computer games can improve your health and fitness," Springer, 2014, pp. 1–7.

[3] Larissa Müller, Sebastian Zagaria, Arne Bernin, Abbes Amira, Naeem Ramzan, Christos Grecos, and Florian Vogt A, "EmotionBike: a study of provoking emotions in cycling exergames," International Conference on Entertainment Computing, Springer, 2015, pp. 155–168.

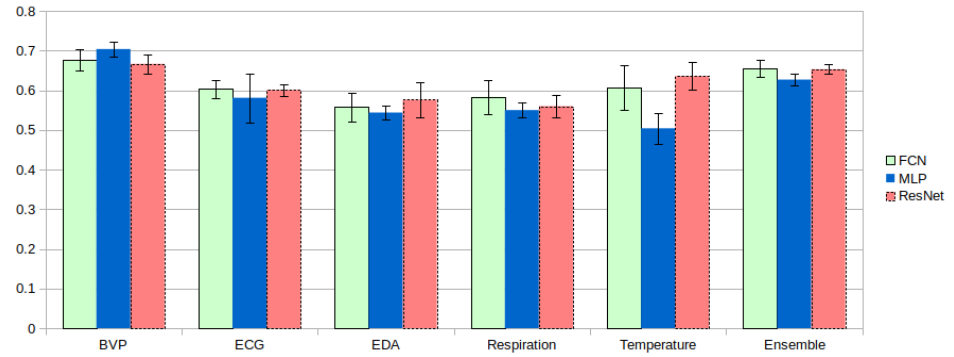

Fig. 5. Results of the Experiments

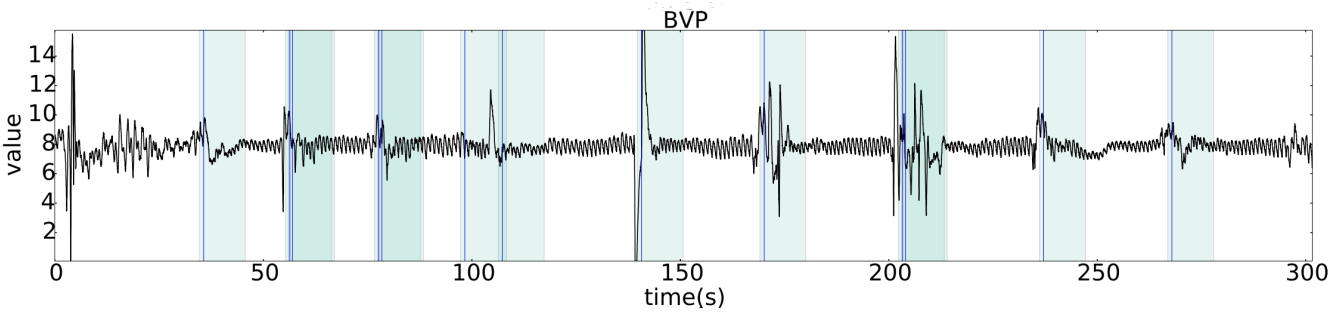

Fig. 6. Typical time series of BVP sensor data in the *Downhill Scene*

[4] Larissa Müller, Arne Bernin, Sobin Ghose, Wojtek Gozdzielewski, Qi Wang, Christos Grecos, Kai von Luck, and Florian Vogt, "Physiological data analysis for an emotional provoking exergame," IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2016, pp. 1–8.

[5] Larissa Müller, Arne Bernin, Kai von Luck, Andreas Kamenz, Sobin Ghose, Qi Wang, Christos Grecos, and Florian Vogt, "Enhancing Exercise Experience with Individual Multi-Emotion Provoking Game Elements," IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2017, pp. 1–8.

[6] Larissa Müller, Arne Bernin, Kai von Luck, Andreas Kamenz, Sobin Ghose, Qi Wang, Christos Grecos, and Florian Vogt, "Emotional Journey for an Emotion Provoking Cycling Exergame," 4th Intl. Conf. on Soft Computing & Machine Intelligence (ISCMI 2017), IEEE, 2017, pp. 104–108.

[7] Tamanna Tabassum Khan Munia et al. "Mental states estimation with the variation of physiological signals," International Conference on Informatics, Electronics & Vision (ICIEV), IEEE, 2012, pp. 800–805.

[8] Avinash Parnandi, Youngpyo Son, and Ricardo Gutierrez-Osuna, "A Control-Theoretic Approach to Adaptive Physiological Games," Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, 2013, pp. 7–12.

[9] Thomas Christy, and Ludmila I. Kuncheva, "Technological advancements in affective gaming: A historical survey," GSTF Journal on Computing (JoC) 3.4, 2014, p. 32.

[10] Neal Lathia et al. "Happier People Live More Active Lives: Using smartphones to link happiness and physical activity," Public Library of Science, 2016.

[11] Stuart J. H. Biddle, Kenneth R. Fox, and Stephen H. Boutcher, "Chapter 4: Emotion, mood and physical activity," Physical Activity and Psychological Well-being. Routledge, Psychology Press, 2000, p. 63.

[12] Darren ER Warburton et al. "The health benefits of interactive video game exercise," Applied Physiology, Nutrition, and Metabolism, Volume 32, NRC Research Press, 2007, pp. 655–663.

[13] Gunnar ER Borg, "Anstrengungsempfinden und körperliche Aktivität," Deutsches Ärzteblatt 101.15, 2004, A1016A1021.

[14] Georgios N Yannakakis, Hector P Martinez, and Maurizio Garbarino, "Psychophysiology in games," Emotion in Games, Springer, 2016, pp. 119–137.

[15] Wolfram Boucsein, " Electrodermal activity," Springer Science & Business Media.

[16] E-H Jang et al. "Identification of the optimal emotion recognition algorithm using physiological signals," International Conference on Engineering and Industries (ICEI), IEEE, 2011, pp. 1–6.

[17] B. Zhong et al., "Emotion recognition with facial expressions and physiological signals," 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, 2017, pp. 1–8.

[18] Mandar Gogate, Adeel Ahsan, and Amir Hussain. "Deep learning driven multimodal fusion for automated deception detection," Computational Intelligence (SSCI), 2017 IEEE Symposium Series on. IEEE, 2017.

[19] Zhiguang Wang, Yan Weizhong, and Tim Oates. "Time series classification from scratch with deep neural networks: A strong baseline," Neural Networks (IJCNN), 2017 International Joint Conference on. IEEE, 2017.

[20] Mahabal, Ashish, et al. "Deep-learnt classification of light curves," Computational Intelligence (SSCI), 2017 IEEE Symposium Series on. IEEE, 2017.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Adv. in Neural Information Processing Systems 25, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 10971105.

[22] Anthony Bagnall, et al. "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," Data Mining and Knowledge Discovery 31.3 (2017): 606-660.

[23] Maximilian Christ, Andreas W. Kempa-Liehr, and Michael Feindt. "Distributed and parallel time series feature extraction for industrial big data applications," arXiv preprint arXiv:1610.07717 (2016).

[24] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep sparse rectifier neural networks," Proceedings of the fourteenth international conference on artificial intelligence and statistics. 2011.

[25] Kaiming He, et al. "Deep residual learning for image recognition," Proc. of the IEEE conf. on computer vision and pattern recognition. 2016.

[26] R. Polikar. "Ensemble Learning," In: Zhang C., Ma Y. (eds) Ensemble Machine Learning. Springer, Boston, MA. 2012.