Article

Comparative Analysis of Emotion Classification Based on Facial Expression and Physiological Signals Using Deep Learning

SeungJun Oh ^{1,2} and Dong-Keun Kim ^{3,4,*}

¹ Department of Kinesiology, University of Maryland, College Park, MD 20742, USA; seungjun0215@gmail.com

² Neuromechanics Research Core, University of Maryland, College Park, MD 20742, USA

³ Department of Human-Centered Artificial Intelligence, Sangmyung University, Seoul 03016, Korea

⁴ Institute of Intelligence Informatics Technology, Sangmyung University, Seoul 03016, Korea

* Correspondence: dkim@smu.ac.kr

Featured Application: Emotion classification based on facial expression and physiological signals.

Abstract: This study aimed to classify emotion based on facial expression and physiological signals using deep learning and to compare the analyzed results. We asked 53 subjects to make facial expressions, expressing four types of emotion. Next, the emotion-inducing video was watched for 1 min, and the physiological signals were obtained. We defined four emotions as positive and negative emotions and designed three types of deep-learning models that can classify emotions. Each model used facial expressions and physiological signals as inputs, and a model in which these two types of input were applied simultaneously was also constructed. The accuracy of the model was 81.54% when physiological signals were used, 99.9% when facial expressions were used, and 86.2% when both were used. Constructing a deep-learning model with only facial expressions showed good performance. The results of this study confirm that the best approach for classifying emotion is using only facial expressions rather than data from multiple inputs. However, this is an opinion presented only in terms of accuracy without considering the computational cost, and it is suggested that physiological signals and multiple inputs be used according to the situation and research purpose.

Keywords: emotion classification; deep learning; physiological signal; facial expression

Citation: Oh, S.; Kim, D.-K.

Comparative Analysis of Emotion Classification Based on Facial Expression and Physiological Signals Using Deep Learning. *Appl. Sci.* **2022**, *12*, 1286. <https://doi.org/10.3390/app12031286>

Academic Editor: Giancarlo Mauri

Received: 22 December 2021

Accepted: 24 January 2022

Published: 26 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human emotions indicate a person's psychological state. There are positive emotions and negative emotions. Positive emotions are related with improving work efficiency as well as human health [1]. Negative emotions due to various factors cause stress and reduce concentration [2]. In addition, the various negative stimuli that people experience in their daily lives are highly diverse, and the accumulation of these stimuli causes mental disorders such as stress and depression [3]. Negative emotions intensify stress, and constant exposure to stress causes serious diseases.

The definition of emotion is presented differently by each study, but in general, it can be divided into two ways. One of them is the discrete emotion model and the other is the dimensional model. In the discrete emotion model, many studies use the six-basic-emotions model of Ekman [4], which consists of happiness, sadness, surprise, anger, fear, and disgust. These can be divided into positive and negative emotions. Positive emotions include happiness and surprise, and negative emotions include sadness, anger, fear, and disgust [5]. In addition, the nine affects of Silvan Tomkins [6,7] and Wheel of Emotions of Plutchik [8] are also included in the discrete emotion model. Unlike the discrete model, which considers all emotions individually, the dimensional model considers all emotions as continuous. It is based on that humans themselves define that their individual emotions exist in two or three dimensions [9]. Most dimensional emotional models consist of an integrated form

of valence, arousal, or intensity of emotion. In general, the two-dimensional circumplex model of Russell [10] and the three-dimensional PAD emotional state model [11] are used.

There are many methods for measuring human emotions. A survey-based measurement is basically used to figure out the emotion. However, since the results of emotions cause physiological reactions in the human body, a hormone-based measurement as well as measurements using a non-invasive method are used. A diagnosis by a psychiatrist is necessary when medical care is required based on the results of the measurement [12]. Representatively, methods such as the Cornell Medical Index (CMI) [13], saliva and urinalysis cortisol test for physiological responses, and hormone-level test through a blood test were generally used. Each measurement method has advantages and disadvantages. For the method using a questionnaire, the error in measurement accuracy cannot be ignored as the answer varies depending on the subjectivity. The hormone-measurement method has the disadvantage that it cannot grasp the status in real time [14]. In contrast, the non-invasive measurement method can complement real-time performance and is considered to be fairly reliable because it is based on physiological signals that the subject cannot change artificially [3].

Conventionally, research on emotion classification using physiological signals and facial expressions has been performed based on simple computational calculations. As these studies gradually develop, methods using machine learning and pattern recognition have been proposed, and more recently, studies to classify emotions using deep learning with computer hardware performance have been proposed. This type of research is widely used not only in emotion recognition [3,15,16], posture recognition [17,18] and medical diagnosis and therapy [19,20], but also in finance [21] and even agriculture [22].

Studies have shown that physiological signals and emotion are closely related. In the study of Kreibig [23], emotion and the autonomic nervous system (ANS) response specificity were reviewed, and Nasoz [24] presented how emotions and cognitive processes are physiologically affected. In addition, a study by Levenson [25], who conducted four different experiments to confirm the correlation between emotion-based facial expressions and physiological signals, was also conducted.

Negative emotions can be expressed as stress and anxiety and are closely related. Stress can be recognized for its cause, but anxiety often involves not recognizing what the object of anxiety is. Stress is a response to the past and present, while anxiety is also different in that it involves worry about what will happen in the future. Many studies have been conducted to estimate stress based on physiological signals [26–28], and recently, stress analysis methods that apply physiological signals to artificial intelligence models have also been developed [29–31]. Research using various physiological signals as input data have yielded highly reliable results. In the study conducted by Wijsman et al. [32], the human emotional status was classified into various types by machine learning after acquiring the ECG, respiration, skin conductance, and EMG signals using a wearable device, and classified almost 80% accuracy. Similarly, in the study conducted by Sandulescu et al. [33], stress was classified using a support vector machine based on photoplethysmography (PPG), electrodermal activity (EDA), and heart rate variability (HRV) parameters, and classified about 79% accuracy and precision. Additionally, Convolution Bidirectional Long Short-term Memory Neural Network (CBLNN) frameworks by Du et al. [34] classify emotions as the geometric features extracted from facial skin information and the heart rate extracted from changes in RGB component. Similar to these studies that analyzed the response of the autonomic nervous system, studies using signals from the central nervous system have also been conducted [14,35–37].

Many studies have classified emotion using such non-invasive signals, and studies using facial expressions for stress analysis are also being conducted [38,39]. Human emotions are not expressed through conscious effort, but spontaneously, and are accompanied by changes in the face itself and physiological changes in facial muscles. In particular, a person's facial expression conveys an emotional state and a feeling and plays an important role in non-verbal communication due to the person's emotions seen in the face itself [40].

In more detail, a study has been conducted to predict depression, anxiety, and stress levels using the facial action coding system (FACS). This study presents 87.2% accuracy for depression, 77.9% for anxiety, and 90.2% for stress [41]. In the study of Busso et al. [42], emotions were classified using both facial expressions and acoustic information, and the performance of both modalities was evaluated. Similarly, there is a study by Hossain et al. [43] that analyzes data with two different modalities. Regarding the multi-modalities, various studies are being conducted [44,45], and the combination of heterogeneous data has great potential because it can be applied to various topics.

Additionally, the emotion classification using deep learning typically focuses on the accuracy itself, and the subject of research is often fixed by simply presenting the analysis results. Therefore, we focused our research on the emotion classification process, and in this study, we classified emotion using facial expressions and physiological signals using deep learning and then compared and analyzed the classification results for all the input data. Furthermore, we presented the results of classification according to input data applied to deep-learning models, based on which we also presented efficient emotion classification methods.

2. Methods

2.1. Experiment

For this study, we recruited a total of 53 adult subjects who had no underlying physical or mental disease and conducted the experiment. All of the participants consisted of adult students or office workers and were set as a group that could fully understand all the conditions and situations suggested during the experiment. The trial was registered in the IRB of experiment institute (IRB approval number: 30-2017-63). Ethical approval was provided by the IRB at the site, and the study adhered to the principles outlined in the Declaration of Helsinki and good clinical practice guidelines. In the experiment, 32 men and 21 women participated, with an average age of 29 ± 10 years.

The subjects were asked to express a total of four emotions—happiness, sadness, anger, and surprise—for 15 s per expression, as shown in the reference picture in Figure 1. In this procedure, the reference picture was provided to the subjects in advance, and the subjects made their facial expressions as imitating the reference picture. Additionally, the subjects' facial expressions were totally recorded. When the experiments acquiring facial expressions ended, four videos showing emotions were watched for 1 min. We defined happiness and surprise as positive emotions and sadness and anger as negative emotions. While the subjects were showing emotional expressions, their facial expressions were recorded using a camera, and the BVP signal was measured while watching the video. The device used in the experiment was a BVP finger-clip sensor of NeXus-10 MKII of Mind Media, and the frame rate of the camera used to record the facial expressions was 30 frames/s. To stimulate the participants, video clips expressing happiness, surprise, anger, and sadness were played, in that order, for 1 min. The corresponding video clips used in the experiment were from the movies *About Time* (2013), *Capricorn One* (1977), *The Attorney* (2013), and *The Champ* (1979). The overall procedure of the experiment is shown in Figure 2.

Figure 1. Reference picture and sample of actual test subject.

Figure 2. Overall procedure of the experiment.

2.2. Dataset

2.2.1. Facial Expression Data

We recorded all the facial expressions of the subjects and acquired their facial images by cropping the part recognized as the face in all the frames of the recorded image. To recognize the faces, the face_recognition library of Python [46] was used. Next, the parts that were judged to be noise in the data, such as the face being covered during the experiment or inappropriate expressions, were deleted. Finally, the preprocessing of the image data was completed through the augmentation of the data via lateral inversion.

The amount of acquired data was from a total of 53 subjects, with an image rate of 30 frames/s, a time duration of 15 s, and two out of four emotions considered as negative emotion and positive emotion, respectively. The number of data points was approximately 47,700, and this number doubled when data augmentation was performed. Finally, the number of images obtained was approximately 190,000 ($53 \times 30 \times 15 \times 2 \times 2$), but only 157,000 facial expression image data were used, excluding inappropriate data through handcrafted filtering. The preprocessing procedure of image data is presented in Figure 3.

Figure 3. Image data preprocessing procedure.

2.2.2. Physiological Signal Data

After recording the facial expressions of the subjects, we made them watch a video that revealed their emotions for 1 min. In addition, physiological signals of the subjects were acquired using a BVP sensor. HRV parameters (heart rate, HRV amplitude, LF, HF, and LF/HF) were extracted from the acquired BVP signal with BioTrace+ of mind media B.V. In the experiment, the physiological data of four subjects were excluded owing to signal instability and other reasons. In other words, the amount of physiological data used in this study was from 49 subjects, with 32 samples/s signal, 1 min time duration, and each of two out of four emotions was defined as negative emotion and positive emotion, respectively. For labeling, the amount of data per each emotion label was approximately 188,000, and approximately 376,000 physiological data were used. Table 1 lists the number

of images and physiological signals. The input data for use in the emotion classification model were prepared.

Table 1. Number of data samples used as input data.

Data	Label	Emotion	<i>n</i>	Frame Rate	Time (s)	Data Augmentation	Amount of Data per Label	Number of Data ¹	Total Amount ²
Facial data	Negative emotion	Ang ³	53	30	15	Double increase	47,700	190,800	157,000
		Sad ⁴					47,700		
	Positive emotion	Hap ⁵					47,700		
		Sur ⁶					47,700		
Physio signal	Negative emotion	Ang ³	49	32	60	-	94,080	376,320	376,000
		Sad ⁴					94,080		
	Positive emotion	Hap ⁵					94,080		
		Sur ⁶					94,080		

¹ Before handcrafted filtering. ² After handcrafted filtering. ³ Anger. ⁴ Sadness. ⁵ Happiness. ⁶ Surprise.

2.3. Emotion Classification Model

To classify emotion according to input data and compare the results, we constructed a deep-learning model to which physiological signals and facial expressions were applied. The model constructed for this study is based on the convolution neural network (CNN). The principle of CNN divides the input image into pixels and find the pattern based on divided pixels to recognize the images. Thus, the feature extract is not required, unlike the conventional machine learning. The neural network constructed on this principle has the advantage of being reliable, fast, and accurate. Additionally, since the CNN model can be used again for a new recognition task based on the existing neural network, various models have been introduced.

The constructed deep-learning model was named a physio- and facial-emotion classification model based on the input data. Based on the accuracy, we compared the classification effects according to the input data of the emotion classification model. The preprocessed data were applied to all the constructed models. We used Python's scikit-learn library, Keras, and Tensorflow to develop a deep learning model, and 70% of the total data were used as a training set and the remaining 30% as a testing set. In addition, a model was built to classify physiological signals by customizing a CNN model. The structure of each model and the parameters used for training are presented in Figure 4 and Table 2, respectively.

Figure 4. Structure of physio-emotion classification model (a) and facial-emotion classification model (b).

Table 2. Hyper parameters applied to physio-emotion classification model and facial-emotion classification model.

Classifier	Physio-Emotion Classification Model	Facial-Emotion Classification Model
Batch size	32	32
Learning rate	0.0001	0.001
Optimizer	Adam Optimizer	
Cost function	Cross Entropy	

3. Result

Before discussing the deep-learning classification results, the classification evaluation metrics suggested in the study are as below, and Equations (1)–(4) are formulas used to calculate the classification evaluation metrics [3,47].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

True positive (TP), false positive (FP), true negative (TN), and false negative (FN).

We classified emotion using the physiological signal (positive, negative) applied to the physio-emotion classification model explained in the methods section. Table 3 presents the confusion matrix of classifying emotion using physiological signals. Table 4 presents the results of classifying emotion by applying physiological signals to the physio-emotion classification model.

Table 3. Confusion matrix of classifying emotion using physiological signals.

	Negative Emotion	Positive Emotion
Negative Emotion	43,592	12,830
Positive Emotion	8003	48,404

Table 4. Results of classifying emotion using physiological signals.

Evaluation Metrics	Accuracy	Precision	Recall	F1 Score
Result	0.8154	0.7905	0.8581	0.8299

We also classified emotion by applying the facial image to the facial-emotion classification model. Table 5 presents the confusion matrix of classifying emotion using facial expressions. Table 6 presents the results of classifying emotion by applying facial images to the facial-emotion classification model. Based on the classification results, it was confirmed that using a facial image is a better classification method when classifying emotion.

Table 5. Confusion matrix of classifying emotion using facial expressions.

	Negative Emotion	Positive Emotion
Negative Emotion	24,836	2
Positive Emotion	1	22,372

Table 6. Results of classifying emotion using facial expressions.

Evaluation Metrics	Accuracy	Precision	Recall	F1 Score
Result	0.9999	0.9999	0.9999	0.9999

4. Discussion

As mentioned earlier, positive emotions improve the work efficiency and provide beneficial effects on mental health. However, negative emotions cause serious stress, and continuous exposure to stress causes various diseases such as depression [3]. Negative emotion occurs in physical, social, and personal environments. It is accompanied by negative psychological and physiological symptoms in the body. In general, we called this stress; in the case of mild stress, symptoms disappear in a short time, but in the case of severe stress, symptoms persist for a long time.

Similar to our proposed method, many deep-learning-based studies [3,15,29–33,43,47] have recently been conducted using data obtained from physiological signals. However, unlike these studies that focus on the results, we focused our research on the process. Thus, we attempted to classify emotion in various ways and suggested a more efficient method of classifying it. To achieve the purpose of our study, we recorded the facial expressions of users revealing emotions and obtained physiological signals based on the experiments. In addition, through the preprocessing of experimental data, an appropriate amount of data was obtained, and a deep-learning model was constructed to classify the emotion of the user. The user's emotion was classified as 81% when the emotion was classified using only physiological signals, and the classification result was confirmed as 99% when emotion was classified using facial-expression images. Consequently, it was confirmed that classifying emotion using facial images is a better method. It can be inferred that the reason that the classification result of facial expressions is higher is probably because the amount of information it contains is larger than the physiological signals due to the characteristics of the image data.

However, the purpose of our study was to focus on the emotion analysis process, and we would like to expand the research results. Therefore, we would like to discuss the classification results by constructing a multimodal classification model that applies physiological signals and facial expression data to a deep-learning model, recognized as a physio-facial-emotion classification model. Based on this result, it was concluded that an emotion classification study using deep learning can be conducted more efficiently. In this classification model, facial expressions and physiological signals were applied as input data to a deep learning model. Therefore, the features of data were extracted from the physio- and facial-emotion classification models created in the Methods section, and the extracted features were concatenated again. The two concatenated features were classified as regressions to calculate the accuracy of the model. The composition of the constructed physio-facial-emotion classification model is shown in Figure 5. The weights, parameters, and structure of this model were the same as those of the single model (the physio- and the facial-emotion classification model). Using the same weights, parameters, and structure, the classification metrics of the concatenated model can be easily compared to those of the single model.

We applied physiological signals and facial expressions to the physio-facial-emotion classification model. Table 7 is confusion matrix of classifying emotion using physiological signals and facial expressions, and the classification results are presented in Table 8.

When we analyzed the results of the unimodal and multimodal classifications, the results of classifying emotion using only facial expressions showed higher accuracy than classification using multiple data. Therefore, when attempting to classify emotion, using only facial images is the most efficient way, and it was confirmed that there is no significant effect even if multiple data are used to improve accuracy. Therefore, we would like to suggest that the method of analyzing emotion by acquiring a facial image rather than a physiological signal is a more efficient method, given an experimental situation in which

both a facial image and a physiological signal can be acquired when conducting an emotion analysis study.

Figure 5. Structure of the physio-facial-emotion classification model.

Table 7. Confusion matrix of classifying emotion using physiological signal and facial expression.

	Negative Emotion	Positive Emotion
Negative Emotion	164,683	24,953
Positive Emotion	26,560	157,230

Table 8. Results of classifying emotion using physiological signal and facial expression.

Evaluation Metrics	Accuracy	Precision	Recall	F1 Score
Result	0.8621	0.8630	0.8555	0.8592

However, this study has some limitations. First, it was necessary to synchronize the two experiments. In our experiment, the subjects' facial expressions were recorded with a posed expression that imitates the standard photographic sample, and then an emotion-stimulation video was viewed. In other words, we did not acquire the spontaneous facial images while watching the video. Indeed, the recognition and accuracy of facial expressions according to spontaneous and posed faces have not yet been researched [48], but in this study, facial expressions should have been recorded simultaneously in the experiment process. The second limitation is the lack of emotions. We classified the remaining four emotions into negative and positive emotions, excluding disgust and fear, among six basic emotions [4]. However, disgust and fear are also clear forms of negative emotion, and if these two emotions are added to the classification model, the classification results are unpredictable. Therefore, in future studies, the remaining two emotions should be included, and other positive emotions should be included. In other words, it is necessary to expand the category of emotion and conduct an emotion classification study.

5. Conclusions

In conclusion, because humans inevitably face various emotions in their lives, it is important to manage it. Various studies have been conducted to manage the emotions including negative emotions. In this study, we not only presented emotion classification results for various input data but also designed a multimodal model to compare and analyze an efficient emotion classification method. Although there are some limitations of this

study, the presented results are fairly original because the study focused on the emotion classification process, unlike other studies that only focused on emotion classification results. We believe that our research results will be of great help to future research related to psychological and emotion research and management, and further research should be conducted to overcome the limitations of this research.

Author Contributions: Conceptualization, S.O. and D.-K.K.; methodology, S.O.; validation, D.-K.K.; data curation, S.O.; writing—original draft preparation, S.O.; writing—review and editing, S.O. and D.-K.K.; supervision, D.-K.K.; project administration, D.-K.K.; funding acquisition, D.-K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (Grant number: NRF-2021R1F1A1063817) and Basic Science Research Program through the NRF grant funded by the Ministry of Education (Grant number: NRF-2021R1A6A3A03044487).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of SMG-SNU Boramae Medical Center (IRB approval number: 30-2017-63. Date of approval: 8 November 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to subjects' privacy.

Acknowledgments: This paper is partially based on Seungjun Oh's Ph.D. dissertation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ali, M.; Mosa, A.H.; Machot, F.A.; Kyamakya, K. Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review. In *Recent Advances in Nonlinear Dynamics and Synchronization. Studies in Systems, Decision and Control*; Kyamakya, K., Mathis, W., Stoop, R., Chedjou, J., Li, Z., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 109, pp. 287–302.
2. Maaoui, C.; Pruski, A. Emotion Recognition through Physiological Signals for Human-Machine Communication. In *Cutting Edge Robotics 2010*; Kordic, V., Ed.; IntechOpen: London, UK, 2010; pp. 317–332.
3. Oh, S.; Lee, J.-Y.; Kim, D.K. The Design of CNN Architectures for Optimal Six Basic Emotion Classification Using Multiple Physiological Signals. *Sensors* **2020**, *20*, 866. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Ekman, P. The argument and evidence about universals in facial expressions of emotion. In *Handbooks of Psychophysiology*; Wagner, H., Manstead, A., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 1989; pp. 143–164.
5. Kanagaraj, G.; Ponnambalam, S.G.; Jawahar, N. Supplier Selection: Reliability Based Total Cost of Ownership Approach Using Cuckoo Search. In *Trends in Intelligent Robotics, Automation, and Manufacturing, Proceedings of the First International Conference, IRAM 2012, Kuala Lumpur, Malaysia, 28–30 November 2012*; Ponnambalam, S.G., Parkkinen, J., Ramanathan, K.C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 491–501.
6. Tomkins, S.S. The Positive Affects. In *Affect Imagery Consciousness: The Complete Edition*; Springer: New York, NY, USA, 2008; pp. 3–285.
7. Tomkins, S.S. Anger and Fear. In *Affect Imagery Consciousness: The Complete Edition*; Springer: New York, NY, USA, 2008; pp. 687–975.
8. Plutchik, R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **2001**, *89*, 344–350. [\[CrossRef\]](#)
9. The American Psychological Association, Dimensional theory of emotion in APA Dictionary of Psychology. Available online: <https://dictionary.apa.org/dimensional-theory-of-emotion> (accessed on 9 May 2021).
10. Russell, J. A Circumplex Model of Affect. *J. Pers. Soc. Psychol.* **1980**, *39*, 1161–1178. [\[CrossRef\]](#)
11. Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Curr. Psychol.* **1996**, *14*, 261–292. [\[CrossRef\]](#)
12. Sharma, N.; Gedeon, T. Objective measures, sensors and computational techniques for stress recognition: A survey. *Comput. Methods Programs Biomed.* **2012**, *108*, 1287–1301. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Abramson, J.H. The cornell medical index as an epidemiological tool. *Am. J. Public Health Nations Health* **1966**, *56*, 287–298. [\[CrossRef\]](#)
14. Kang, J.S.; Jang, G.; Lee, M. Stress status classification based on EEG signals. *JIIBC* **2016**, *16*, 103–108. [\[CrossRef\]](#)

15. Oh, S.; Kim, D.K. Development of the Stacked ensemble model for convenient emotion classification using respiratory signals. In Proceedings of the 2019 International Conference and Exhibition on Computational Biology and Bioinformatics (ICECBB 2019), Taipei, Taiwan, 1–2 December 2019; Wiley: Hoboken, NJ, USA, 2019; p. 21.
16. Jerrieta, S.; Murugappan, M.; Nagarajan, R.; Wan, K. Physiological signals based human emotion recognition: A review. In Proceedings of the 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, Malaysia, 4–6 March 2011; IEEE: New York, NY, USA, 2011; pp. 410–415.
17. Oh, S.; Kim, D.K. Development of Squat Posture Guidance System Using Kinect and Wii Balance Board. *JICCE* **2019**, *17*, 74–83.
18. Oh, S.; Kim, D.K. Posture Classification Model Based on Machine Learning for Guiding of Squat Exercises. In Proceedings of the 2019 IERI International Conference on Medical Physics, Medical Engineering and Informatics (ICMMI 2019), Tokyo, Japan, 22–24 March 2019; Wiley: Hoboken, NJ, USA, 2019; pp. 359–360.
19. Wang, T.; Xuan, P.; Liu, Z.; Zhang, T. Assistant diagnosis with Chinese electronic medical records based on CNN and BiLSTM with phrase-level and word-level attentions. *BMC Bioinform.* **2020**, *21*, 1–16. [\[CrossRef\]](#)
20. Bharatharaj, J.; Huang, L.; Al-Jumaily, A.M.; Krageloh, C.; Elara, M.R. Experimental evaluation of parrot-inspired robot and adapted model-rival method for teaching children with autism. In Proceedings of the 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), Phuket, Thailand, 13–15 November 2016; IEEE: New York, NY, USA, 2016; pp. 1–6.
21. Barra, S.; Carta, S.M.; Corrigan, A.; Podda, A.S.; Recupero, D.R. Deep learning and time series-to-image encoding for financial forecasting. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 683–692. [\[CrossRef\]](#)
22. Narvekar, C.; Rao, M. Flower classification using CNN and transfer learning in CNN-Agriculture Perspective. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 3–5 December 2020; IEEE: New York, NY, USA, 2020; pp. 660–664.
23. Kreibitz, S.D. Autonomic nervous system activity in emotion: A review. *Biol. Psychol.* **2010**, *84*, 394–421. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Nasoz, F.; Alvarez, K.; Lisetti, C.L.; Finkelstein, N. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognit. Technol. Work.* **2004**, *6*, 4–14. [\[CrossRef\]](#)
25. Levenson, R.W.; Ekman, P.; Friesen, W.V. Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology* **1990**, *27*, 363–384. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Thayer, J.F.; Ahs, F.; Fredrikson, M.; Sollers, J.J.; Wager, T.D. A meta-analysis of heart rate variability and neuroimaging studies. *Neurosci. Biobehav. Rev.* **2012**, *36*, 747–756. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Steptoe, A.; Marmot, M. Impaired cardiovascular recovery following stress predicts 3-year increases in blood pressure. *J. Hypertens.* **2005**, *23*, 529–536. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Pedrotti, M.; Mirzaei, M.A.; Tedesco, A.; Chardonnet, J.R.; Merienne, F.; Benedetto, S.; Baccino, T. Automatic stress classification with pupil diameter analysis. *Int. J. Hum.-Comput. Interact.* **2014**, *30*, 220–236. [\[CrossRef\]](#)
29. Song, S.H.; Kim, D.K. Development of a Stress Classification Model Using Deep Belief Networks for Stress Monitoring. *Healthc. Inform. Res.* **2017**, *23*, 285–292. [\[CrossRef\]](#)
30. Li, R.; Liu, Z. Stress detection using deep neural networks. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 285. [\[CrossRef\]](#)
31. Bobade, P.; Vani, M. Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. In Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 15–17 July 2020; IEEE: New York, NY, USA, 2020.
32. Wijsman, J.; Grundlehner, B.; Liu, H.; Hermens, H.; Penders, J. Towards Mental Stress Detection using wearable physiological sensors. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; IEEE: New York, NY, USA, 2011.
33. Sandulescu, V.; Andrews, S.; Ellis, D.; Bellotto, N.; Mozos, O.M. Stress Detection Using Wearable Physiological Sensors. In *Artificial Computation in Biology and Medicine, Proceedings of the International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC) 2015, Elche, Spain, 1–5 June 2015*; Springer: Cham, Switzerland, 2015.
34. Du, G.; Wang, Z.; Gao, B.; Mumtaz, S.; Abualnaja, K.M.; Du, C. A Convolution Bidirectional Long Short-Term Memory Neural Network for Driver Emotion Recognition. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4570–4578. [\[CrossRef\]](#)
35. Asif, A.; Majid, M.; Anwar, S.M. Human stress classification using EEG signals in response to music tracks. *Comput. Biol. Med.* **2019**, *107*, 182–196. [\[CrossRef\]](#)
36. Reza, A.S.; Setaredan, S.K.; Nasrabadi, A.M. Classification of mental stress levels by analyzing fNIRS signal using linear and non-linear features. *Int. Clin. Neurosci. J.* **2018**, *5*, 55–61.
37. Blanco, J.A.; Vanleer, A.C.; Calibo, T.K.; Firebaugh, S.L. Single-Trial Cognitive Stress Classification Using Portable Wireless Electroencephalography. *Sensors* **2019**, *19*, 499. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Gao, H.; Yüce, A.; Thiran, J. Detecting emotional stress from facial expressions for driving safety. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP 2014), Paris, France, 27–30 October 2014; IEEE: New York, NY, USA, 2014.
39. Zhang, H.; Feng, L.; Li, N.; Jin, Z.; Cao, L. Video-Based Stress Detection through Deep Learning. *Sensors* **2020**, *20*, 5552. [\[CrossRef\]](#)
40. Ali, M.F.; Khatun, M.; Turzo, N.A. Facial Emotion Detection Using Neural Network. *IJSER* **2020**, *11*, 1318–1325.
41. Gavrilescu, M.; Vizireanu, N. Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System. *Sensors* **2019**, *19*, 3693. [\[CrossRef\]](#) [\[PubMed\]](#)

42. Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, C.M.; Kazemzadeh, A.; Lee, S.; Neumann, U.; Narayanan, S. Analysis of emotion recognition using facial expressions, speech and multimodal information. In Proceedings of the ICMI '04: 6th international conference on Multimodal interfaces, State College, PA, USA, 13–15 October 2004; Association for Computing Machinery: New York, NY, USA, 2004; pp. 205–211.
43. Hossain, M.S.; Muhammad, G. Emotion recognition using deep learning approach from audio–visual emotional big data. *Inf. Fusion* **2019**, *49*, 69–78. [[CrossRef](#)]
44. Bayoudh, K.; Knani, R.; Hamdaoui, F.; Mtibaa, A. A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets. *Vis. Comput.* **2021**, 1–32. [[CrossRef](#)]
45. Romaissa, B.D.; Mourad, O.; Brahim, N. Vision-Based Multi-Modal Framework for Action Recognition. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; IEEE: New York, NY, USA, 2021.
46. Face Recognition, GitHub. Available online: https://github.com/ageitgey/face_recognition (accessed on 22 January 2021).
47. Oh, S.; Kim, D.K. Machine–Deep–Ensemble Learning Model for Classifying Cybersickness Caused by Virtual Reality Immersion. *Cyberpsychol. Behav. Soc. Netw.* **2021**, *24*, 729–736. [[CrossRef](#)]
48. Eom, J.S.; Oh, H.S.; Park, M.S.; Sohn, J.H. Discrimination between spontaneous and posed smile: Humans versus computers. *Sci. Emot. Sensib.* **2013**, *16*, 95–106.