Machine Learning Models and Videos of Facial Regions for Estimating Heart Rate: A Review on Patents, Datasets, and Literature

Tiago Palma Pagano [1,†], **Victor Rocha Santos** [1,†], **Yasmin da Silva Bonfim** [2,†], **José Vinícius Dantas Paranhos** [3,†], **Lucas Lemos Ortega** [4,†], **Paulo Henrique Miranda Sá** [4,†], **Lian Filipe Santana Nascimento** [5,†], **Ingrid Winkler** [6,†] and **Erick Giovani Sperandio Nascimento** [6,†,*]

1   Computational Modeling Department, SENAI CIMATEC University Center, Salvador 41650010, Brazil; tiago.pagano@fbter.org.br (T.P.P.); victorocha11@gmail.com (V.R.S.)
2   Artificial Intelligence Technician, SENAI CIMATEC University Center, Salvador 41650010, Brazil; yasmin.bonfim@fbter.org.br
3   Data Science, Estácio de Sá University, Salvador 41098020, Brazil; jose.paranhos@fbter.org.br
4   Computer Engineering, SENAI CIMATEC University Center, Salvador 41650010, Brazil; lucas.ortega@fieb.org.br (L.L.O.); paulo.sa@fbter.org.br (P.H.M.S.)
5   Information Systems, Federal University of Bahia, Salvador 41650010, Brazil; lian.nascimento@fbter.org.br
6   Department of Management and Industrial Technology, SENAI CIMATEC University Center, Salvador 41650010, Brazil; Ingrid.winkler@doc.senaicimatec.edu.br
*   Correspondence: erick.sperandio@fieb.org.br
†   These authors contributed equally to this work.

Abstract: Estimating heart rate is important for monitoring users in various situations. Estimates based on facial videos are increasingly being researched because they allow the monitoring of cardiac information in a non-invasive way and because the devices are simpler, as they require only cameras that capture the user's face. From these videos of the user's face, machine learning can estimate heart rate. This study investigates the benefits and challenges of using machine learning models to estimate heart rate from facial videos through patents, datasets, and article review. We have searched the Derwent Innovation, IEEE Xplore, Scopus, and Web of Science knowledge bases and identified seven patent filings, eleven datasets, and twenty articles on heart rate, photoplethysmography, or electrocardiogram data. In terms of patents, we note the advantages of inventions related to heart rate estimation, as described by the authors. In terms of datasets, we have discovered that most of them are for academic purposes and with different signs and annotations that allow coverage for subjects other than heartbeat estimation. In terms of articles, we have discovered techniques, such as extracting regions of interest for heart rate reading and using video magnification for small motion extraction, and models, such as EVM-CNN and VGG-16, that extract the observed individual's heart rate, the best regions of interest for signal extraction, and ways to process them.

Keywords: heart rate; region of interest; facial image; machine learning

1. Introduction

The heart rate is a vital human body signal that allows monitoring of a person's health. Heart rate estimation is particularly important for monitoring users in a range of frequent situations, such as driving a vehicle [1], engaging in physical activities [2], working in hazardous conditions [3], and during investigative police interviews [4]. Heart rate or its variability may be used to track and detect stress-related aspects [1], tiredness [5], emotions [6], health [7], and social behavior [8].

Heart rate is the number of times the heart beats blood in one minute [9]. In general, heart rate ranges from 41 to 240 beats per minute (BPM) [10,11]. For a resting adult, heart rate can range from 60 to 100 BPM [9,11,12]. Electrocardiogram (ECG) is a test that measures

the resting heartbeat rhythm [13] and can be used to diagnose the patient's heart health. Photoplethysmography (PPG), on the other hand, is a technique for measuring blood volume variations [14], which are commonly obtained by a wrist or finger oximeter.

Real-time heart rate monitoring allows one to observe a patient's health, as it is linked to many different factors in the body. Remote heart rate estimation is very important because of the high cost of other types of estimators, and it is also a less invasive technique because it does not require physical contact and is low cost [5,15,16], reducing the risk of possible infection, for instance in premature babies [17]. In addition, some applications require long-term monitoring, and the use of contact devices for a longer period may cause discomfort or irritation [12]. Remote estimation allows physiological information [18] to be obtained without the need for physical contact, making it possible to observe a patient's health [18] and help choose the best treatment, while being simpler than physical options.

Visual estimation, which is often conducted with a camera and is less costly and less intrusive to the user, is an alternative to traditional measurement approaches [16]. As a result, research into heart rate estimation through video has increased, as has the usage of pp. in different heart rate assessment devices.

Machine learning (ML) generates a process of learning from a model, simulating human learning by undergoing training to enable computers to learn real-world concepts and their relationships through the experiences contained in the data, with the goal of solving or assisting in solving problems. From the data are extracted several factors with characteristics that influence the outcome. These variations cannot always be easily observed in the data, but they influence the results [19]. ML has been used for many different purposes, such as 3D object generation, drug creation, pandemic studies, image processing, face detection, image to text translation, texture transfer, traffic control, noise removal in images [20], and even heart rate measurement.

Deep learning (DL) has been used to estimate heart rate from facial recordings [10,12,21,22]. DL techniques are multi-layered artificial neural networks (ANN) with great degrees of flexibility, which enables the efficient and effective classification of a wide variety of circumstances. Using a dataset, the generalization of problems is achievable by modifying the internal parameters of the network to reflect the desired structure. This enables the discovery of optimum combinations of complex feature data. This particular ability enables the development of autonomous systems capable of making human-like decisions.

Deep convolutional neural networks (DCNNs) are used to learn relevant patterns from a representation and estimate heart rate. For instance, [10,21] use the VGG network, while [12] uses a ResNet-18, and [22] proposes a VGG style network with 3D-CNN layers to extract features from videos over time. As can be seen, and as [12] exposes, there are two categories used to build a DL framework to estimate heart rate: end-to-end and feature-decoder methods. Commonly, feature-decoder methods remove non-skin regions [10], select regions of interest [10,12,21], denoise [10], build spatiotemporal features [10,12,21], and perform other preprocessing tasks [22].

There has been no prior study that has reviewed the benefits and constraints of using machine learning models to estimate heart rate from facial videos, to the best of the authors' knowledge. Nonetheless, this comprehension might play a vital role in the development of applications in this sector. Furthermore, patents give strategic information to industry and academics, in addition to being a resource for technical management and innovation. As a consequence, patent examination may specify a particular technology, in this way allowing for the discovery of technology advancements, inventors, market trends, and other features.

Several virtual reality and augmented reality devices are already equipped with cameras that capture the user's face; these can be leveraged for heart rate estimation and monitoring of the user's condition in certain contexts [12].

Thus, this study aims to examine existing knowledge to analyze the benefits and challenges of using machine learning models to estimate heart rate from facial videos through patents, datasets, and article reviews.

This paper is organized into four parts. Section 2 outlines the methods used, Section 3 details our findings, and Section 4 presents our final considerations and recommendations for future investigations.

2. Materials and Methods

This systematic patent, dataset, and literature review adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards, which have been developed to "help systematic reviewers transparently report why the review was done, what the authors did, and what they found" [23]. In addition, the method described in [24] was used, which comprises seven stages: planning, defining the scope, searching the literature, assessing the evidence base, synthesizing, analyzing, and writing. To assess risk of bias in the included studies, as per PRISMA item 5 [23], the preliminary search strategy was designed by a team of five machine learning (ML) model researchers. Then, the candidate strategy was peer-reviewed by three senior ML researchers. The developed method is explained in the following sections.

2.1. Planning

The knowledge bases that will be examined are determined during the planning stage [24]. We chose the Derwent Innovation Index database for patent search because of its vast coverage: it comprises 39.4 million patent families and 81.1 million pieces of patent data, as well as 59 foreign patent agencies and two journal sources. The Derwent platform includes distinct features that increase data extraction, such as the "Smart Search" feature, which uses artificial intelligence to boost keyword finding [25].

Another tool is the Derwent World Patents Index (DWPI), the world's most extensive database of enhanced patent information, which includes extended patent titles and abstracts, English abstracts of the original patent, and an advanced categorization system [26].

We used Google's indexer to perform an exploratory search of face video datasets to be used for heart rate estimation since there are no dedicated databases for searching datasets.

Scopus, Web of Science, and IEEE Xplore were used to search for articles. They were chosen because they are worldwide multidisciplinary scientific databases with extensive citation indexing coverage. Scopus now includes 81 million inspected documents, Web of Science has over 82 million items, and IEEE Xplore has approximately 5 million documents.

2.2. Defining the Scope

The stage of defining the scope is focused on well-stated research questions [24]. We determined three key research questions to investigate, which are as follows:

Q1: How are patents on using machine learning models to estimate heart rate from facial videos characterized in terms of assignees, publications per year, and advantages of the invention?

Q2: How are the available datasets in ANN training for heart rate estimation characterized?

Q3: How is the current knowledge on the use of machine learning models to estimate heart rate from facial videos characterized, in terms of use cases, methodologies and models, and comparison of metrics?

2.3. Searching the Literature

This stage comprises scanning the databases indicated in the planning stage for a phrase relevant to the research questions specified in the stage of defining the scope [24].

We searched through the databases using the following potential search phrase:

 (((("heart rate" OR "heart activity" OR "heart rate estimation") AND ("vital signs" OR "remote ppg" OR "remote photoplethysmography") AND ("Image" OR "Video") AND ("face" OR "facial")))

This potential search phrase was then fed into the Biblioshiny tool for bibliometrics analysis. As shown in Figure 1, the co-occurrence network of terms revealed "PPG", "heart", and "heart rate" as the most key elements of the two clusters created.
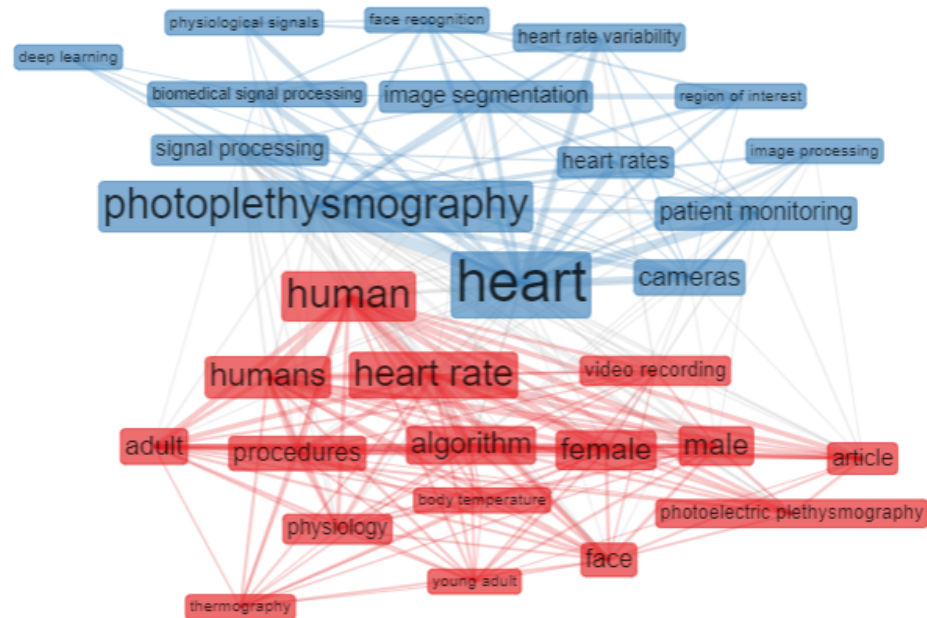
Figure 1. Keyword co-occurrence network.

2.4. Assessing the Evidence Base

The preliminary search was adjusted at this stage, so that inclusion and exclusion criteria were established to determine whether the evidence was eligible or ineligible, as per PRISMA item 5 [23]. Additionally, the LitSearch library was adopted to optimize the proposed initial search phrase, as well as to minimize the risk of bias of the publications included.

("blood volume" OR "health monitoring" OR "signal processing" OR "vital sign") AND ("facial region" OR "facial expression" OR "imaging photoplethysmography" OR "facial video") AND ("heart" OR "volume pulse" OR "heart rate") AND ("convolutional neural network" OR "deep learning" OR "neural network")

We limited the amount of records returned at the previous stage by applying the following exclusion criteria:

E1: Exclude patents filed or articles published before 2016;
E2: Exclude items that are not written in English;
E3: Exclude dead patent applications.

Similar search keywords were used to obtain articles and datasets, with minor changes to meet the search engine parameters of the Scopus, Web of Science, and IEEE Xplore databases.

The search was conducted in May 2021, and we screened seven patent records, eleven datasets, and twenty articles after applying exclusion criteria.

2.5. Synthesizing and Analyzing

Derwent analytical and "Insights" tools were used to scrutinize the patents. The recovered items were exported to the Mendeley Reference Manager tool, and Microsoft Excel visuals were developed.

Figure 2 represents the stages of this systematic review.
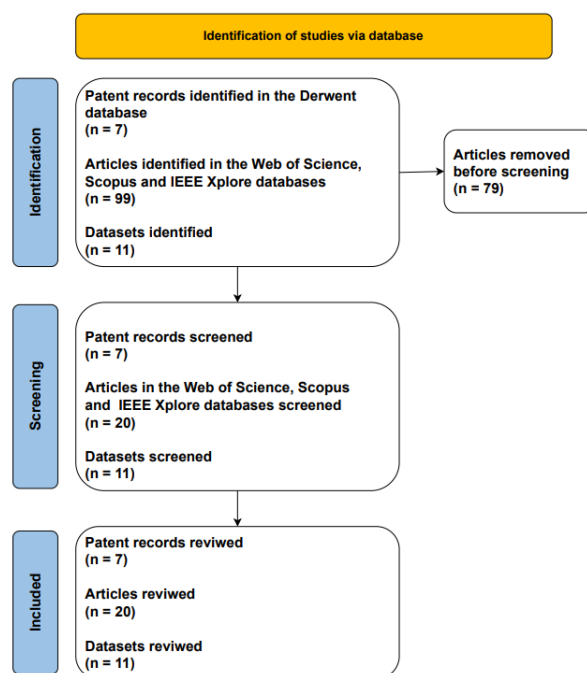
Figure 2. Systematic review flow diagram, adapted from PRISMA 2020.

3. Results And Discussion

The research questions Q1, Q2, and Q3 were examined in order to discover the benefits and limitations of employing ML for heart rate estimation from facial video. In the sections that follow, we analyze our findings.

3.1. Patent landscape

Table 1 shows the seven patent records identified.

To address the first research question, these patents were analyzed to answer frequent concerns and uncover patterns in assignees, publications per year, and benefits.

Q1: How are patents on using machine learning models to estimate heart rate from facial videos characterized in terms of assignees, publications per year, and advantages of the invention?

The characterization of patent assignees may contribute to the identification of industry leaders, the assessment of potential competitors, and the identification of niche players. Except for GB2572961A, we discovered that the majority of published patents are from Chinese applicants. Furthermore, the widely dispersed distribution of published patents among multiple assignees is interesting: just one assignee, Hangzhou First People's Hospital, submitted two patents, whilst the other patents were filed by distinct assignees.

Instead of a relatively equal-sized but vast portfolio held by a few organizations—indicating an active competitive market with strong investments by multiple companies, thus implying that the market is difficult to enter—we discovered a large number of assignees, each with a small number of records, which indicates a space for developing technology. Acquisition or quick development may be used to enter this area. There are multiple companies, each with a low number of patents, which signals a chance to approach this field while it is still in its inception, either by licensing existing technology, acquiring one of the competitors, or inventing new technology that is not currently patented. It is also important to note that three of the seven assignees are from the health sector, and three of the seven assignees are universities.

We found a substantial increase in annual patent publication in 2020. There were no patents published until 2018, then one was published in 2019, and six were published in

2020. Because of the 18-month patent confidentiality rule, we excluded 2021 patents from our analysis.

In terms of benefits, we examined the seven patents in relation to the advantages of the invention as described by the authors and the novelty of the invention, that is, the unique innovative feature introduced by the inventor that is not conventional and that is an improvement on existing technology.

Table 1. List of retrieved patents.

Identification Number	Title	Assignee
WO2019202305A1	System for vital sign detection from a video stream	ClinicCo Ltd. (London, UK)
CN110738155A	Face recognition method and device, computer equipment and storage medium	Hangzhou First People's Hospital (Hangzhou, China)
CN110909717A	Moving object vital sign detection method	Nanjing University of Science and Technology (Nanjing, China)
CN111259787A	Unlocking method and device, computer equipment and storage medium	Hangzhou First People's Hospital (Hangzhou, China)
CN111260634A	Facial blood flow distribution extraction method and system	Tianjin Polytechnic University (Tianjin, CN)
CN111797794A	Facial dynamic blood flow distribution detection method	People's Public Security University of China (Bejing, China)
US20200155040A1	Systems and methods for determining subject positioning and vital signs	Hill Rom Services Inc. (Batesville, US)

Patent US2020155040-A1 proposes a method for automatically monitoring a patient's position and vital signs by using near-infrared (NIR) cameras and long wavelength Infrared (LWIR) to monitor hospital patients by measuring heart rate and facial temperature.

Patent CN111797794-A proposes a method to detect the flow of blood distribution on a person's face. A video of the face is captured with a red green blue (RGB) camera to determine a region of interest (ROI), the ROI heart rate pattern is determined via remote photoplethysmography (rPPG), and blood flow is obtained from each sub-region. The value of blood flow intensity is obtained from the pulse wave of each sub-region. In addition, the distribution of facial blood flow can be obtained from the value of the blood flow intensity identified in each sub-region. This process does not require subject skin contact to be tested and is highly accurate.

Patent CN110909717-A aims to detect the vital signs of a moving person by performing face recognition, as well as obtaining the heart rate by displaying the information in real time. Vital signs are extracted from the subject's face and detected with the help of face detection; however, the detection of vital signs can fail. In addition, heart rate values can be obtained by vital signs, and the caching and displaying of the heart rate value is done in real time. This method reduces the complexity of the measurement and improves the accuracy of the signal extraction.

A method for extracting facial blood flow distribution is proposed by patent CN111260-634-A. The video of the person's face is captured with an RGB camera, and each frame of the video is divided into sub-regions; in this way, pulse rate is determined in each separate sub-region using an rpp. algorithm. The pulse wave signal of the sub-regions of a frame is obtained by rPPG, and its signal is used as the value of the blood flow distribution, thus allowing the identification of facial blood flow distribution. In addition, heart rate frequency is determined by the maximum amplitude in a spectrogram.

Patent CN111259787-A proposes a method to unlock doors or devices with heart rate and state of inebriation based on data from a person's face captured by a camera, aiming to enhance biometric security. Face recognition, heart rate detection, and inebriation detection are performed based on face information. In addition, human face image and human face

temperature information are also obtained to aid in heart rate and inebriation detection. Face recognition and heart rate detection are performed based on the face image, which allows the obtaining of results for face recognition and heart rate detection.

Patent CN110738155-A proposes a method of facial recognition that includes capturing the head and the entire body, applying image amplification with Eulerian video magnification (EVM) to extract the original signal and heart rate. Human face information and human heart rate information are obtained based on face image and off-face image. This method avoids the influence of external light and body detection, resulting in an accurate answer for face recognition. In addition, human heart rate information is obtained based on the rate information environment.

Finally, patent WO2019202305-A1 proposes a method to detect a user's heart rate from transmitted video. The face is detected in the frames, and its motion is analyzed using face detection. Heart rate is determined by changing the colors of pixels and face movement in the video motion. Additionally, the heart rate of the subject is determined based on the selection of one of the first and second estimated pulse frequencies.

3.2. Datasets

Table 2 shows the eleven datasets selected by the search strategy.

Table 2. List of retrieved datasets.

Datasets	Number of Videos	Subjects	Presence of Infrared Videos
MMSE-HR	102	40	No
VIPL-HR	3230	107	Yes
Deap—Part 01	120	14–16	No
Deap—Part 02	40	32	No
COHFACE	160	40	No
MAHNOB	3741	27	Yes
ECG-Fitness Dataset	207	17	No
r-pp.	21	03	No
MR-NIRP	180	18	Yes
Imaging Photoplethysmography Dataset	60	12	No
Toadstool	10	10	No
UBFC-Rpp.	42	42	No

These eleven datasets were analyzed to address the second research question:

Q2: How are the available datasets in ANN training for heart rate estimation characterized?

We found out that most of the datasets serve academic objectives and come with a variety of licenses.

The MMSE-HR (Multimodal Spontaneous Expression-Heart Rate) dataset [27] is composed of facial emotions annotated in terms of occurrence and intensity. In addition, each video's associated heart rate and blood pressure sequences are included in the dataset. It comes in three different licensing types: standard, individual, and corporate. Because of the large number of participants, the dataset may have a benefit over the others in terms of heart rate estimate.

The VIPL-HR dataset [28] includes videos of persons in nine different situations of head movement and lighting, as well as their heart rates. The dataset is exclusively accessible for academic study and needs the completion of a release form. As recorded in Table 2, the dataset has the advantage of incorporating infrared facial capture and the largest number of individuals among the examined datasets.

The DEAP dataset [29] is an emotion analysis dataset that employs electroencephalogram, physiological, and visual signals. It is broken into two sections. In the first one, the participants rate segments of music videos on arousal, valence, and dominance. In the second experiment, the participants watch and rate the same movies as before while their electroencephalogram and physiological data are recorded. In addition, RGB videos of the participants' faces are captured. The dataset is only permitted for use in academic research, and form submission is needed for the study's purposes.

The COHFACE dataset [30] contains RGB videos of persons' faces in 480p and 20 frames per second, as well as their blood volume and breathing rate pulses. It is necessary to fill out a form with justification in order to obtain access to it. The EULA license covers the dataset.

The MAHNOB dataset [31] features RGB videos of persons' faces in six different poses, as well as their ECG, electroencephalogram (EEG), breath amplitude, and skin temperature data. The dataset is protected under the EULA license and is available for download on the website.

The ECG-Fitness dataset [32] includes 1080p RGB films of persons doing physical exercises on fitness equipment, as well as their ECG signals. It is free for academic usage and requires the signing of a compliance agreement. By presenting six situations among its videos, the dataset includes a greater variety of heart rates.

The r-pp. Algorithm performance dataset [33] is used to assess the robustness of rpp. against changes and high fluctuations in heart rate and pulse rate. It includes RGB videos of subjects as well as their distinct ECG signals. The dataset is free to download and is licensed under the 4TU general terms of use. In comparison with previous datasets, R-pp. Algorithm has fewer participants and no infrared video, which can be observed in Table 2.

The MERL-Rice Near-Infrared Pulse (MR-NIRP) [34] is a publicly available dataset that is separated into two contexts: within a vehicle and inside a room. The MR-NIRP Vehicle has films of the faces of individuals in a car, for a total of ten experiments per person. The MR-NIRP Indoor is recorded in a room, with each of the eight participants' faces recorded either static or moving. The contexts include files containing data from a pulse oximeter connected to the participants' fingers. Sixteen of the participants are males, four of them have beards, and two are women. They are all between the ages of 20 and 60, with Asian, Indian, and Caucasian skin tones. The only requirement to use the datasets is to reference them in publications.

The Imaging Photoplethysmography Dataset [35] is a public dataset that contains RGB videos (1920 × 1200) of the faces of participants as well as their PPGs and BPMs. The dataset is freely accessible.

The Toadstool [36] dataset is intended for academic usage only, although it may be used commercially with a license subscription. It is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license and includes 480p videos of the faces of participants playing Super Mario Bros, as well as heart rate data, blood volume pulse, and beat intervals extracted from an Empatica E4 wristband.

The dataset Univ. Bourgogne Franche-Comté Remote Photoplethysmography (UBFC-RPPG) [37] comprises 480p videos of the participants' faces and their individual pulse signals.

As previously stated, there is a vast range of datasets accessible, each with unique participant distribution and information architecture.

3.3. Article Landscape

Table 3 lists the twenty articles that were selected.

These articles were reviewed in order to answer research question Q3, and our findings are provided in the subsections that follow:

Q3: How is the current knowledge on the use of machine learning models to estimate heart rate from facial videos characterized, in terms of use cases, methodologies and models, and comparison of metrics?

3.3.1. Bibliometric Analysis

Figure 1 depicts the total extent of the key terms found in the reviewed articles. The terms are clearly divided into two groups: the blue group refers to terms that we will use to define the technical scope of the bibliometric analysis, while the red group identifies the research objectives associated with the technical terms.

Figure 3 depicts the extracted word cloud from the reviewed articles, indicating which terms are most related to the subject. The terms with the greatest incidence numbers are connected to cardiac monitoring using cameras that display a human face image. Other terminology used to describe comparable approaches include image segmentation, face recognition, image and signal processing, independent component analysis, and others.

Table 3. List of retrieved articles.

Ref.	Title	Publication Year
[6]	Emotion recognition from facial expressions and contactless heart rate using knowledge graph.	2020
[36]	Toadstool: A dataset for training emotional intelligent machines playing Super Mario Bros.	2020
[10]	A deep learning framework for heart rate estimation from facial videos.	2020
[1]	Non-contact-based driver's cognitive load classification using physiological and vehicular parameters.	2020
[9]	Non-Contact Emotion Recognition Combining Heart Rate and Facial Expression for Interactive Gaming Environments.	2020
[16]	Visual Heart Rate Estimation from Facial Video Based on CNN.	2020
[5]	On assessing driver awareness of situational criticalities: Multi-modal bio-sensing and vision-based analysis, evaluations, and insights.	2020
[22]	DeepPerfusion: Camera-based Blood Volume Pulse Extraction Using a 3D Convolutional Neural Network.	2020
[12]	Heart Rate Estimation from Facial Videos Using a Spatiotemporal Representation with Convolutional Neural Networks.	2020
[38]	Robust remote heart rate estimation from face utilizing spatial-temporal attention.	2019
[39]	Automatic Monitoring of Driver's Physiological Parameters Based on Microarray Camera.	2019
[17]	Combating the impact of video compression on non-contact vital sign measurement using supervised learning.	2019
[40]	Architectural tricks for deep learning in remote photoplethysmography.	2019
[41]	Emotion inference of game users with heart rate wristbands and artificial neural networks.	2019
[15]	EVM-CNN: Real-Time Contactless Heart Rate Estimation from Facial Video.	2019
[11]	Long Distance Vital Signs Monitoring with Person Identification for Smart Home Solutions.	2018
[18]	A Novel Short-Term Event Extraction Algorithm for Biomedical Signals.	2018
[21]	Deep learning with time-frequency representation for pulse estimation from facial videos.	2018
[42]	Deep super resolution for recovering physiological information from videos.	2018
[43]	Towards Generic Modelling of Viewer Interest Using Facial Expression and Heart Rate Features.	2018
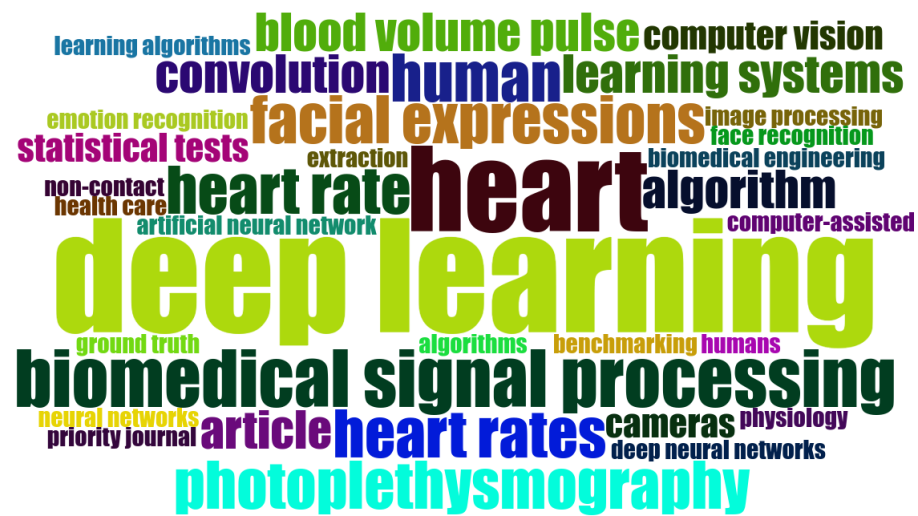
Figure 3. Keyword search cloud.

The conceptual structure of the reviewed articles is shown in Figure 4. This structure covers almost 97% of the primary subjects mentioned in the articles. The clusters represent how ideas are connected; the closer they are, the stronger the association. There is a cluster of topics about techniques and another about applications. Figure 4 demonstrates the density that the technical terms in Figure 1 have in relation to the terms that refer to the object of study.
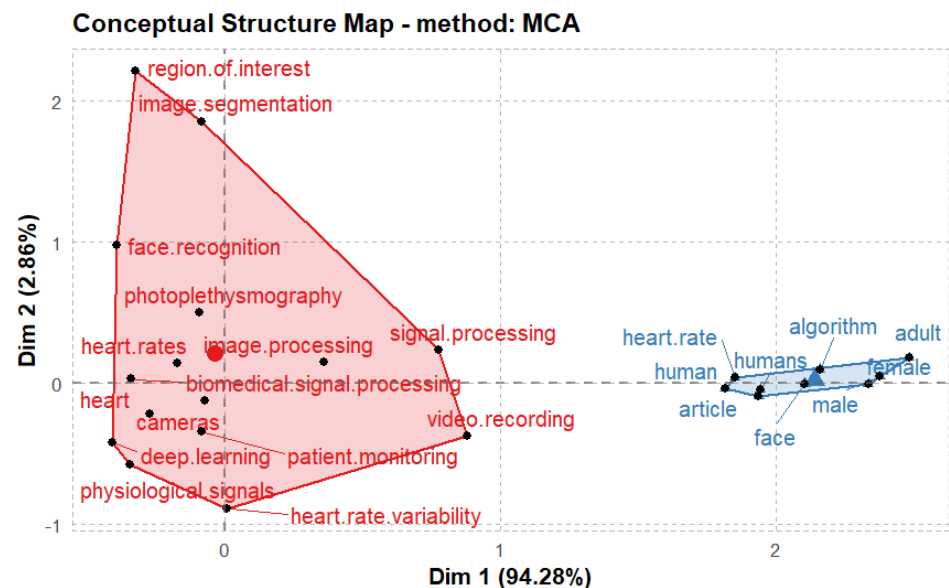
Figure 4. Conceptual framework map.

Figure 5 illustrates the writers who had the biggest impact on the research subject. Among them, the authors Ambikapathi, A. and McDuff, D. stand out as having the greatest relevance among those selected for this review. It can be seen that although McDuff, D. shows the highest relevance in Figure 5, Ambikapathi, A. shows the largest name in Figure 6, and therefore, shows a higher impact in the thematic group.
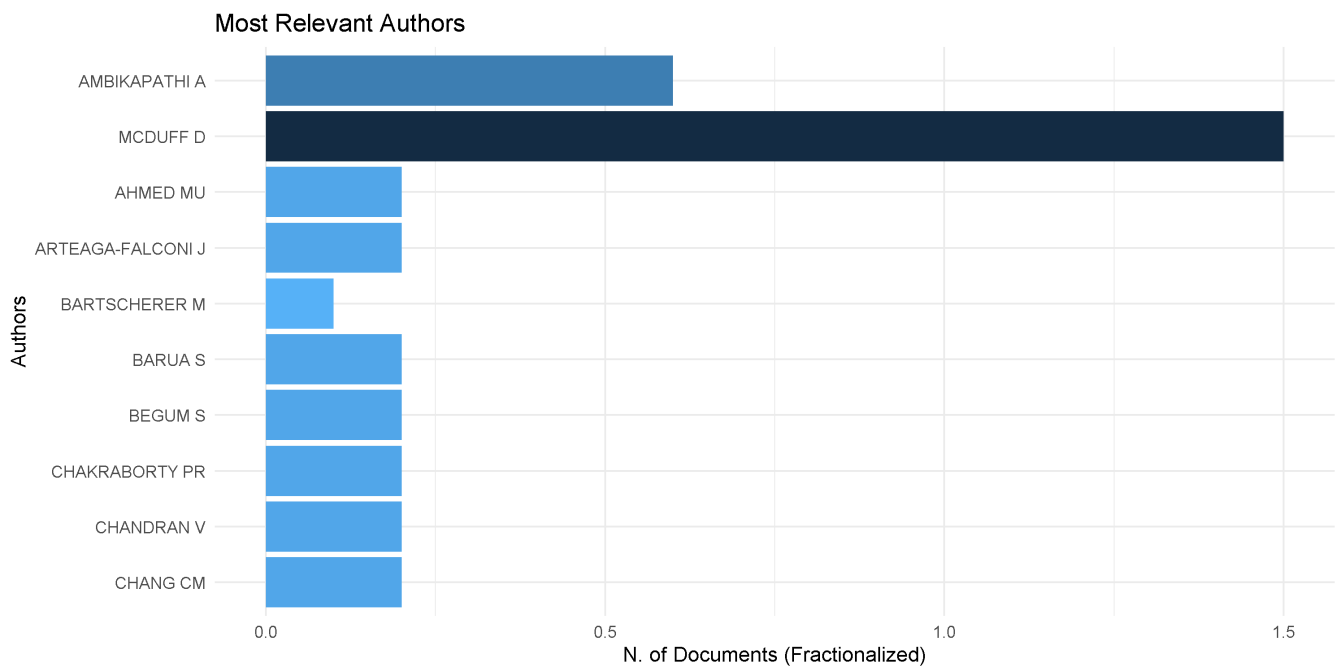
Figure 5. Impact of authors.

Figure 6 shows the collaborative networks of authors found in the publications; in addition, the longer the author's name, the greater their appearance in the works found. The division by color reveals the group of authors who worked on the same work and the size of the node indicates the impact on the thematic group of the research.
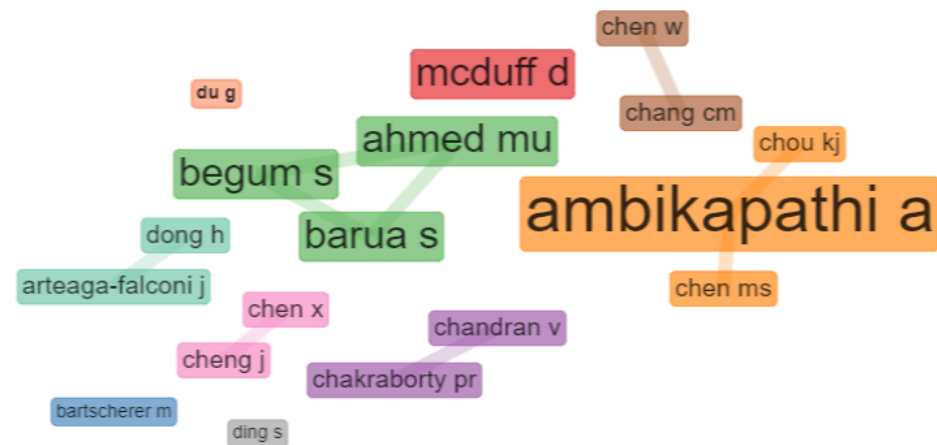
Figure 6. Collaboration network.

3.3.2. Use Cases

In [6,10–12,15,18,21,39,40], the authors have reduced the ROIs of the face using facial landmark detectors, while [9,16,17,22,38,42] have used the full face as model input. In [10,21], the authors have removed of irrelevant signals such as tremor, illumination changes, and so on in order to locate and cut out ROIs to estimate heart rate. In [15], it was demonstrated that using either the whole face or ROIs offers positive results for heart rate estimation, but both have advantages and disadvantages. When the whole face is used, there is extra processing to clean up the signals. When there is ROI clipping, [15] has pointed out that, despite requiring less processing in terms of treating irrelevant regions, the provision of a consistent detection of face landmarks to smooth noisy signals was a challenge. In [6,9–12,15–18,21,22,38–40,42], heart rate was estimated using face videos,

while [1,5,9,41,43] have estimated additional physiological signals using contact devices or heart rate estimation techniques.

Using case studies, several previously investigated methodologies, as well as their benefits and disadvantages, are identified, as well as the approaches they have taken, as can be seen in Figure 7. This enables the search for alternative answers to further challenges without the need to repeat prior mistakes or commit time to researching a path already taken.
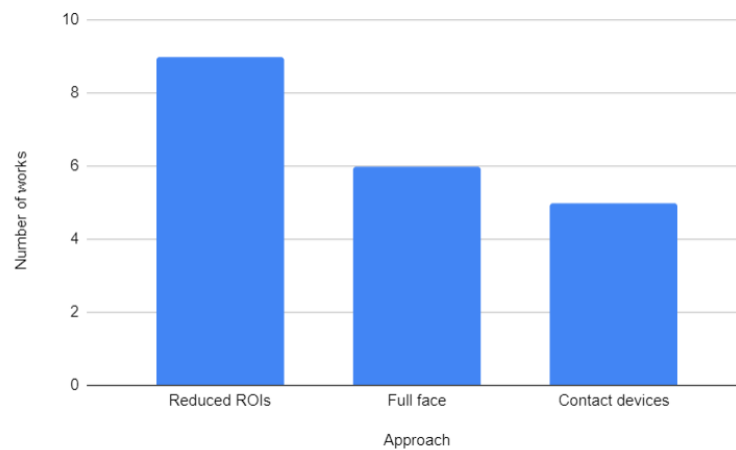
Figure 7. Works' approaches.

3.3.3. Models

There is a wide variety of adopted techniques used for data capture, data processing, and final estimation. Most of the articles have used convolutional neural networks (CNNs) to build ANNs for noise treatment in order to perform heart rate estimations. In [9,16], temporal heart rate information was collected using recurrent layers for long short-term memory (LSTM). Furthermore, some works have employed other techniques to obtain heart rate [11,18,39]. For face detection of ROIs, in [22,40] have employed Opencv, while [10,21] have used a single shot multibox detector (SSD) network. ROIs are used to monitor color changes at places where heart rate estimate is achieved by viewing such changes. EVM was employed by [11] as an amplifier of the raw signal, which was then filtered with the Butterworth bandpass filter (0.67–4 Hz), a filter used by [18,22] that positions the heart rate found in the optimal ranges for analysis. Heart rate was extracted from videos acquired by RGBs cameras, the most popular and widely available on the market, in [9,12,18,21,42]. Independent component analysis (ICA) was employed by [6,39,42] have to allow an independent signal analysis, and [18] has used the relative energy (Rel-En) algorithm, which determines the heart rate given the preprocessed signal from the ROI's green channel.

In [39], the joint approximate diagonalization estimation of real signals (JADER) algorithm and ICA were used to separate the RGB components and determine heart rate. As a result, most studies have preprocessed the extracted signals, mostly for heart rate normalization, while [16,17] have not undertaken preprocessing of the face videos.

Among the multiple methodologies used to validate ANN estimations, in [6,11,12,38] have evaluated heart rate estimation outcomes using mean error calculation metrics such as root mean square error (RMSE), mean absolute error (MAE), mean error rate (HR$_{mer}$), and L1 loss. Furthermore, some works have used the Pearson correlation coefficient to support the assessment of the estimated results by comparing the estimated data to the ground truth data [12,38]. Standard deviation (SD$_e$) was also examined by [6,12,38] to analyze the dispersion of the estimations.

The preferred use of the green channel in the studies of [10,21,39] and the filtering of external influences to the recordings are two key and recurring themes in several articles. The former is because it is more dependable when it comes to extracting physiological

signs from the individual in the video. The latter, on the other hand, is more detrimental to estimation since it employs techniques such as joint blind source separation (JBSS), feature-decoders, and Gaussian filters. The strategy of approaching heartbeat estimation as a classification problem in [40] resulted in improved network performance in estimating heart rate from videos with dynamic circumstances. In contrast, it was discovered in [5] that heart rate variability (HRV) is an excellent metric for classifying cognitive states and is also more robust than heart rate.

3.3.4. Comparison of Metrics

The MAHNOB-HCI dataset was used by [21] during testing. According to the authors, the study surpasses other works in the RMSE, SD_e, and HR_{mer} metrics, but the mean error (M_e) and Pearson correlation metrics yielded similar results. The MAHNOB-HCI dataset was used by [10] during testing, and it was stated that the technique outperformed [21] in the RMSE, Pearson correlation, and SD_e metrics, and it was in the top three in the others. During testing with the VIPL-HR dataset, and according to [10], the technique surpassed previous studies in terms of RMSE, while placing in the top three in the other metrics. Their dataset was employed by [16]. The strategy outperformed the competition in the Pearson product moment correlation (PPMC) metric and was in the top three in the SD_e and RMSE metrics. With the exception of PPMC, EVM-CNN [15] outperformed all other approaches in this study. On all criteria, ref. [15] has outperformed the others using the MMSE-HR dataset. Using the MAHNOB-HCI dataset, the study has obtained the same performance metrics as the others, outperforming them on all criteria.

When the findings from the MAHNOB-HCI dataset of [15] were compared to those of [10,21], in [15] exceeded all measures except RMSE by 0.18 points.

During testing with the [44] dataset, in [17] produced the following findings: (1) the greater the video's compression rate factor (CRF), the higher the MAE and the lower the signal-to-noise ratio (SNR), that is, the worse the quality of heart rate estimate; (2) networks trained and tested with the same CRF had better metrics as compared to results from training with a CRF lower than the test CRF.

During testing with the dataset of [44], in [42] proved that the super-resolution pre-processing network with the image photoplethysmography (iPPG) ICA outperformed standard upsampling methods on all measures except SNR.

Using the dataset they collected, in [11] demonstrates the usage of the EVM approach with the forehead region as ROI and compares the results of the algorithm using the entire face region as ROI. A decline in heart rate estimate performance was indicated by [11] when the entire face was used, and it was also proven that the EVM algorithm is a preprocessing tool that provides quality to heart rate estimation since its absence generated a larger RMSE.

The goal of [9]'s study was to detect participants' emotions using estimated heart rate and other features extracted from a face video. The authors have reported that the algorithm developed to estimate heart rate had an error of six BPM during testing using their dataset. Furthermore, they did not give any comparison of the findings of their algorithm to estimate heart rate.

In [12], the proposed method's results were evaluated twice: once using triple cross-validation on the test samples of the MANHOB-HCI dataset and once using cross-validation between datasets, using the VIPL-HR and UBFC-Rpp. datasets for training and actual MANHOB-HCI samples for testing.

Two experiments were carried out in [38]. In the first test, just the VIPL-HR dataset was used for training and testing. In comparison to the other models, the suggested model outperformed them all, with the exception of M_e and Pearson correlation, where it came in second. VIPL-HR was used for training and MMSE-HR was used for testing in the second test, where it was the best dataset.

During tests using a dataset collected by them, alternative networks were proposed in [40], which were employed in various conditions, such as no movement of the subject's face, movement of the subject's face, or capturing video from cameras. When the face was

not moving, the combined loss (CL) model achieved a higher MAE and cover accuracy. When the subject's face moved, the CL+F model with filtering layers achieved higher MAE and cover values than the other models. Furthermore, in testing with Cam_1 and Cam_2, the CL+F model scored the best in terms of metrics. For Cam_3, the CL model achieved the best MAE and cover results. The CL+F model produced the best results for MAE and cover in the tests conducted with the whole dataset. The estimations of the proposed model were compared to those of a medical device for heartbeat extraction in [39], and the authors found that accuracy was high.

In [6], the proposed model outperformed the ICA and independent vector analysis (IVA) approaches in heart rate estimation, outperforming the ICA and IVA methods employed in the comparison.

Table 4 shows the metrics mentioned by the reviewed studies.

Table 4. Table with the metrics used by the articles.

Metric	Refs.
Mean Error	[10,15,21,38]
Standard Deviation	[6,10,12,15,16,21,38]
Root Mean Squared Error	[6,10–12,15,16,21,38,42]
Mean Error Rate	[6,10,12,15,21,38]
Pearson Correlation	[10,12,15,21,38,42]
Mean Absolute Error	[12,16,17,38,40,42]
Mean Absolute Percentage Error (MAPE)	[16]
Pearson Product Moment Correlation	[16]
Signal-to-Noise Ratio	[17,42]
Coverage at ±3 bpm	[40]

We have found that mean error, standard deviation, root mean squared error, mean error rate, Pearson correlation, and mean absolute error are the metrics most often mentioned by the reviewed articles, while many contain combinations of the metrics assessed. We have also observed that datasets are used differently and various approaches, such as super-resolution, Eulerian video magnification, and combined loss, are used to enhance heart rate estimation.

4. Conclusions

Our review has examined techniques and models that extract the recorded individual's heart rate, the face's best ROIs for extraction of heart rate signals, and the methods for processing, as well as the datasets relevant to the training of these models. We have observed that the majority of the datasets are for academic purposes and require approval from the authors. Furthermore, the vast majority of them provide RGB videos of the participant's face, as well as heart rate, PPG, or ECG data. We have also discovered that using DL models in conjunction with approaches for removing noise from videos improves performance for heart rate estimation. Regression metrics such as mean error, standard deviation, RMSE, MAE, HR_{mer}, and Pearson correlation have been identified to validate the models' performance.

There are some limitations inherent to the method used to search for the papers, including the fact that the dataset searchers were limited to Google Scholar, and also the exclusion of articles and patents published before 2016, and that we only considered papers in English.

We also suggest heart rate estimation from non-speech and moving facial images, as well as the identification of a larger amount of emotions and other vital signs, such as respiratory rate and heart rate variability, as directions for future research.

Author Contributions: Conceptualization, T.P.P., L.L.O., V.R.S., Y.d.S.B., J.V.D.P., P.H.M.S., L.F.S.N., I.W. and E.G.S.N.; methodology, T.P.P., L.L.O., V.R.S., Y.d.S.B. and J.V.D.P.; validation, T.P.P., I.W. and E.G.S.N.; formal analysis, I.W. and E.G.S.N.; investigation, T.P.P., L.L.O., V.R.S., Y.d.S.B. and J.V.D.P.; data curation, T.P.P., I.W. and E.G.S.N.; writing—original draft preparation, T.P.P., L.L.O., V.R.S., Y.d.S.B., J.V.D.P., I.W. and E.G.S.N.; writing—review and editing, T.P.P., L.L.O., V.R.S., Y.d.S.B., J.V.D.P., P.H.M.S., L.F.S.N., I.W. and E.G.S.N.; visualization, T.P.P., L.L.O., V.R.S., Y.d.S.B., J.V.D.P., P.H.M.S., L.F.S.N., I.W. and E.G.S.N.; supervision, T.P.P., V.R.S., I.W. and E.G.S.N.; project administration, I.W. and E.G.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: Research funded by HP Inc. Brazil to be entitled to the financial credit defined in the Art. 4º of Law number 8.248, by 1991 (Computer Law).

Acknowledgments: We gratefully acknowledge the support of SENAI CIMATEC AI Reference Center for the scientific and technical support and the SENAI CIMATEC Supercomputing Center for Industrial Innovation. The authors would like to thank the National Council for Scientific and Technological Development (CNPq) for financial support. Ingrid Winkler is a CNPq technological development fellow (Proc. 308783/2020-4).

Conflicts of Interest: There are no conflict of interest associated with this publication.

References

1. Rahman, H.; Ahmed, M.U.; Barua, S.; Begum, S. Non-contact-based driver's cognitive load classification using physiological and vehicular parameters. *Biomed. Signal Process. Control* **2020**, *55*, 101634. [CrossRef]
2. Schneider, C.; Hanakam, F.; Wiewelhove, T.; Döweling, A.; Kellmann, M.; Meyer, T.; Pfeiffer, M.; Ferrauti, A. Heart rate monitoring in team sports—A conceptual framework for contextualizing heart rate measures for training and recovery prescription. *Front. Physiol.* **2018**, *9*, 639. [CrossRef] [PubMed]
3. Sharma, R.; Goel, D.; Srivastav, M.; Dhasmana, R. Differences in Heart Rate and Galvanic Skin Response among Nurses Working in Critical and Non-Critical Care Units. *J. Clin. Diagn. Res.* **2018**, *12*, CC09–CC12. [CrossRef]
4. Bertilsson, J.; Niehorster, D.C.; Fredriksson, P.; Dahl, M.; Granér, S.; Fredriksson, O.; Mårtensson, J.; Magnusson, M.; Fransson, P.A.; Nyström, M. Towards systematic and objective evaluation of police officer performance in stressful situations. *Police Pract. Res.* **2020**, *21*, 655–669. [CrossRef]
5. Siddharth, S.; Trivedi, M.M. On Assessing Driver Awareness of Situational Criticalities: Multi-modal Bio-Sensing and Vision-Based Analysis, Evaluations, and Insights. *Brain Sci.* **2020**, *10*, 46. [CrossRef]
6. Yu, W.; Ding, S.; Yue, Z.; Yang, S. Emotion Recognition from Facial Expressions and Contactless Heart Rate Using Knowledge Graph. In Proceedings of the 2020 IEEE International Conference on Knowledge Graph (ICKG), Nanjing, China, 9–11 August 2020; pp. 64–69. [CrossRef]
7. Young, H.A.; Benton, D. Heart-rate variability: A biomarker to study the influence of nutrition on physiological and psychological health? *Behav. Pharmacol.* **2018**, *29*, 140. [CrossRef]
8. Colasante, T.; Malti, T. Resting heart rate, guilt, and sympathy: A developmental psychophysiological study of physical aggression. *Psychophysiology* **2017**, *54*, 1770–1781. [CrossRef]
9. Du, G.; Long, S.; Yuan, H. Non-Contact Emotion Recognition Combining Heart Rate and Facial Expression for Interactive Gaming Environments. *IEEE Access* **2020**, *8*, 11896–11906. [CrossRef]
10. Hsu, G.S.J.; Xie, R.C.; Ambikapathi, A.; Chou, K.J. A deep learning framework for heart rate estimation from facial videos. *Neurocomputing* **2020**, *417*, 155–166. [CrossRef]
11. Szankin, M.; Kwasniewska, A.; Sirlapu, T.; Wang, M.; Ruminski, J.; Nicolas, R.; Bartscherer, M. Long Distance Vital Signs Monitoring with Person Identification for Smart Home Solutions. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; Volume 2018, pp. 1558–1561. [CrossRef]
12. Song, R.; Zhang, S.; Li, C.; Zhang, Y.; Cheng, J.; Chen, X. Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 7411–7421. [CrossRef]
13. Martis, R.J.; Acharya, U.R.; Adeli, H. Current methods in electrocardiogram characterization. *Comput. Biol. Med.* **2014**, *48*, 133–149. [CrossRef] [PubMed]
14. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1–R39. [CrossRef] [PubMed]
15. Qiu, Y.; Liu, Y.; Arteaga-Falconi, J.; Dong, H.; Saddik, A.E. EVM-CNN: Real-Time Contactless Heart Rate Estimation From Facial Video. *IEEE Trans. Multimed.* **2019**, *21*, 1778–1787. [CrossRef]
16. Huang, B.; Chang, C.M.; Lin, C.L.; Chen, W.; Juang, C.F.; Wu, X. Visual Heart Rate Estimation from Facial Video Based on CNN. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 9–13 November 2020; pp. 1658–1662. [CrossRef]

17. Nowara, E.; McDuff, D. Combating the Impact of Video Compression on Non-Contact Vital Sign Measurement Using Supervised Learning. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 1706–1712. [CrossRef]

18. Yazdani, S.; Fallet, S.; Vesin, J.M. A Novel Short-Term Event Extraction Algorithm for Biomedical Signals. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 754–762. [CrossRef]

19. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1.

20. Aggarwal, A.; Mittal, M.; Battineni, G. Generative adversarial network: An overview of theory and applications. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100004. [CrossRef]

21. Hsu, G.; Ambikapathi, A.; Chen, M. Deep learning with time-frequency representation for pulse estimation from facial videos. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 383–389. [CrossRef]

22. Scherpf, M.; Ernst, H.; Malberg, H.; Schmidt, M. DeepPerfusion: Camera-based Blood Volume Pulse Extraction Using a 3D Convolutional Neural Network. In Proceedings of the 2020 Computing in Cardiology, Rimini, Italy, 13–16 September 2020; pp. 1–4. [CrossRef]

23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *88*, 105906.

24. Booth, A.; Sutton, A.; Papaioannou, D. *Systematic Approaches to a Successful Literature Review*; SAGE Publications Ltd.: Thousand Oaks, CA, USA, 2016.

25. Grames, E.M.; Stillman, A.N.; Tingley, M.W.; Elphick, C.S. An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods Ecol. Evol.* **2019**, *10*, 1645–1654. [CrossRef]

26. Codes, E.M. Derwent World Patents Index (DWPI). 2018. Available online: https://www.jaici.or.jp/newstn/pdf/dwpi_database_information.pdf (accessed on 27 March 2022).

27. Zhang, Z.; Girard, J.; Wu, Y.; Zhang, X.; Liu, P.; Ciftci, U.; Canavan, S.; Reale, M.; Horowitz, A.; Yang, H.; et al. Multimodal spontaneous emotion corpus for human behavior analysis. In Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3438–3446. Available online: https://openaccess.thecvf.com/content_cvpr_2016/papers/Zhang_Multimodal_Spontaneous_Emotion_CVPR_2016_paper.pdf (accessed on 27 March 2022).

28. Niu, X.; Shan, S.; Han, H.; Chen, X. RhythmNet: End-to-End Heart Rate Estimation From Face via Spatial-Temporal Representation. *IEEE Trans. Image Process.* **2020**, *29*, 2409–2423. [CrossRef]

29. Koelstra, S.; Mühl, C.; Soleymani, M.; Lee, J.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [CrossRef]

30. Heusch, G., Anjos, A.; Marcel, S. A reproducible study on remote heart rate measurement. *arXiv* **2017**, arXiv:1709.00962.

31. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [CrossRef]

32. Spetlik, R.; Cech, J.; Franc, V.; Matas, J. Visual Heart Rate Estimation with Convolutional Neural Network. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 3–6 September 2018.

33. Hoffman, W.; Lakens, D. *Public Benchmark Dataset for Testing rPPG Algorithm Performance*; Technical Report; 4TU.Centre for Research Data: Delft, The Netherlands, 2019.

34. Nowara, E.M.; Marks, T.K.; Mansour, H.; Veeraraghavan, A. Near-Infrared Imaging Photoplethysmography During Driving. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 3589–3600. [CrossRef]

35. Pai, A.; Veeraraghavan, A.; Sabharwal, A. HRVCam: Robust camera-based measurement of heart rate variability. *J. Biomed. Opt.* **2021**, *26*, 022707. [CrossRef] [PubMed]

36. Svoren, H.; Thambawita, V.; Halvorsen, P.; Jakobsen, P.; Garcia-Ceja, E.; Noori, F.; Hammer, H.; Lux, M.; Riegler, M.; Hicks, S. Toadstool: A dataset for training emotional intelligent machines playing Super Mario Bros. In Proceedings of the 11th ACM Multimedia Systems Conference, Istanbul, Turkey, 8–11 June 2020; pp. 309–314. [CrossRef]

37. Bobbia, S.; Macwan, R.; Benezeth, Y.; Mansouri, A.; Dubois, J. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit. Lett.* **2019**, *124*, 82–90. [CrossRef]

38. Niu, X.; Zhao, X.; Han, H.; Das, A.; Dantcheva, A.; Shan, S.; Chen, X. Robust remote heart rate estimation from face utilizing spatial-temporal attention. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8.

39. Zou, J.; Li, Z.; Yan, P. Automatic Monitoring of Driver's Physiological Parameters Based on Microarray Camera. In Proceedings of the 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), Okinawa, Japan, 31 May–3 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 15–18.

40. Kopeliovich, M.; Mironenko, Y.; Petrushan, M. Architectural Tricks for Deep Learning in Remote Photoplethysmography. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 28 October 2019; pp. 1688–1696.

41. Hsiao, C.C.; Zheng, W.D.; Lee, R.G.; Lin, R. Emotion Inference of Game Users with Heart Rate Wristbands and Artificial Neural Networks. In Proceedings of the 2018 International Symposium on Computer, Consumer and Control (IS3C), Taichung, Taiwan, 6–8 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 326–329.

42. McDuff, D. Deep Super Resolution for Recovering Physiological Information from Videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1448–14487. [CrossRef]

43. Chakraborty, P.R.; Tjondronegoro, D.W.; Zhang, L.; Chandran, V. Towards Generic Modelling of Viewer Interest Using Facial Expression and Heart Rate Features. *IEEE Access* **2018**, *6*, 62490–62502. [CrossRef]

44. Estepp, J.R.; Blackford, E.B.; Meier, C.M. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 5–8 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1462–1469.