*electronics*

MDPI

*Article*

# The Real-Time Image Sequences-Based Stress Assessment Vision System for Mental Health

**Mavlonbek Khomidov [1], Deokwoo Lee [1], Chang-Hyun Kim [2,*] and Jong-Ha Lee [3,*]**

[1] Department of Computer Engineering, Keimyung University, Daegu 42601, Republic of Korea; mavlonbek@stu.kmu.ac.kr (M.K.); dwoolee@kmu.ac.kr (D.L.)
[2] Department of Neurosurgery, Keimyung University Dongsan Hospital, Daegu 42601, Republic of Korea
[3] Department of Biomedical Engineering, Keimyung University, Daegu 42601, Republic of Korea
* Correspondence: ppori2k@dsmc.or.kr (C.-H.K.); segeberg@kmu.ac.kr (J.-H.L.)

**Abstract:** Early detection and prevention of stress is crucial because stress affects our vital signs like heart rate, blood pressure, skin temperature, respiratory rate, and heart rate variability. There are different ways to determine stress using different devices, such as the electrocardiogram (ECG), electrodermal activity (EDA), the electroencephalogram (EEG), photoplethysmography (PPG), or a questionnaire-based method of stress assessment. In this study, we proposed a camera-based real-time stress detection system using remote photoplethysmography (rPPG). We trained different machine learning models using three datasets: the SWELL dataset, the PPG sensor dataset, and the last ECG and EEG-based stress dataset. The models with the highest predictive accuracy were used to classify stress based on HR and HRV features obtained from the face using a camera. HR and HRV estimations from the face were validated on the PURE public dataset and the custom dataset. In this study, it was observed that the random forest algorithm performs significantly better than other models, achieving an impressive 99% predictive accuracy in the SWELL dataset. In the second dataset, the logistic regression technique shows the best result, achieving an accuracy rate of 84.24%. In the last dataset, the ensemble model achieved an accuracy rate of 67%. We also checked the proposed algorithm in the process of public speaking to estimate stress in a real-time situation.

**Keywords:** stress detection; HRV; remote photoplethysmography (rPPG); machine learning

check for updates

**Citation:** Khomidov, M.; Lee, D.; Kim, C.-H.; Lee, J.-H. The Real-Time Image Sequences-Based Stress Assessment Vision System for Mental Health. *Electronics* **2024**, *13*, 2180. https://doi.org/10.3390/electronics13112180

Academic Editor: Luca Mesin

Received: 7 May 2024
Revised: 26 May 2024
Accepted: 31 May 2024
Published: 3 June 2024

**Copyright:** © 2024 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

Stress plays an important role in our overall health. However, many people do not take stress seriously, and most people are unaware of their current stress levels. Stress is a trigger for many diseases. Numerous studies have demonstrated that stress can negatively affect individuals with coronary artery disease and increase the risk of stroke [1]. Stress also causes blood pressure to rise, and rapid stress can eventually lead to hypotension. Managing stress is crucial for maintaining overall health and can be effective for reducing blood pressure and the development of hypertension [2]. Other studies have indicated psychological stress, alcohol abuse, clinical infection, trauma, and surgery as some of the main causes of stroke [3]. If stress is not managed in time, it can have severe consequences for professionals in specific fields, such as surgery, aviation, and driving. Therefore, many researchers try to find the stress levels of people [4]. Stress management is crucial for maintaining overall health. By taking stress seriously and employing measures to manage it, individuals can significantly reduce their risk of developing stress-related diseases. Stress can be defined in different ways using different mental stressors, including computer work tasks, the Stroop color and word task, arithmetic tasks, public speech tasks, and academic examinations [5]. To detect stress, researchers often rely on contact methods and devices, such as an electrocardiogram (ECG), electrodermal activity (EDA), and an electroencephalogram, which measures the electrical activity of the brain signals (EEGs).

By pre-processing these signals, it becomes possible to identify specific characteristics that correspond closely with a person's emotional states. Posada-Quintero et al. [6] measured stress in divers by measuring changes in their sweat levels using EDA during water immersion. EDA data were collected from 14 subjects while divers performed a specific Stroop task underwater. Tzevelekakis et al. [7] classify three levels of stress: low, moderate, and high, using ultra-short-term raw ECG signals. They used the DriveDB dataset, in which ECG signals were recorded while drivers were driving. By employing convolutional neural networks (CNNs), they achieved accuracies of 83.55% and 98.77% for the 3-level and 2-level stress classifications, respectively. Some researchers used multiple signals; Keshan et al. [8] collected signals from contact-based devices—ECG, electromyogram (EMG), foot and hand galvanic skin response, and respiration rate—from 17 participants. They detected the stress levels of the drivers' driving periods and divided them into low, moderate, and high stress levels based on traffic conditions. Using all device data, they achieved 97.4% for high stress, using only ECG signals. The proposed system achieved 88.24% accuracy in predicting low, moderate, and high stress levels. This shows that the information obtained through the ECG signal is very useful in determining stress. Heart rate variability (HRV) is increasingly recognized as a powerful and reliable indicator of stress [5]. This variability refers to the variation in the time interval between consecutive heartbeats, which is affected by various physiological and psychological factors, including stress levels. HRV is measured by analyzing the time series of beat-to-beat intervals from heart rate data, providing a non-invasive window into the autonomic nervous system's dynamics. Changes in the nervous system during periods of mental stress can significantly affect heart rate variability (HRV) features. Studies have shown that both long-term HRV analysis from 24-h recordings and ultra-short-term HRV analysis, shorter than 5 min, can detect stress [5,9,10]. Moreover, stress-induced changes in the nervous system can also influence other HRV features, such as the power spectral density, which provides insights into the balance between sympathetic and parasympathetic activity, and various time-domain and frequency-domain features. During periods of stress, significant changes are observed in the time domain features of HRV, specifically, the RR intervals—the time intervals between successive heartbeats—along with the root mean square of successive differences (RMSSD) and pNN50, which measures the number of pairs of successive NN intervals that differ by more than 50 ms. All features decreased during stress. Furthermore, in the frequency domain, the high-frequency (HF) component of HRV, also decreases during stress. Conversely, the low-frequency/high-frequency (LF/HF) ratio and the low-frequency (LF) component increase during stress [11–16]. A higher HRV is generally associated with a healthy, resilient cardiovascular system and a strong ability to adapt to stress. Conversely, reduced HRV suggests a predominance of stress responses, less flexibility in responding to environmental demands, and potentially greater risk for cardiovascular and other stress-related disorders [1–3].

Typically, HRV measurements derived from ECG are conducted in a clinical setting, and operated by professionals with specialized knowledge in the field. However, this approach necessitates a visit to the clinic, which might not always be convenient for continuous monitoring. Alternatively, PPG-based wearable technology offers a practical solution for those looking to measure their HRV outside of a clinical environment. There are a variety of wearable devices available on the market, designed to track HRV, among other health metrics. These gadgets, which can range from smartwatches to fitness trackers, provide the advantage of continuous monitoring in real time, allowing users to keep track of their HRV data throughout the day during various activities. Recent advancements in technology have led to the development of non-contact methods for monitoring physiological signals, among which is the remote photoplethysmography (rPPG) technique. The rPPG technology enables the detection of blood volume changes in the facial skin through a camera that captures light reflected from the skin. In recent studies, researchers have made significant progress and achieved promising results in determining HRV by utilizing camera-based technologies. Huang et al. [17] used the rPPG signal to estimate

HRV from facial videos using chrominance (CHROM)-based methods. To enhance the accuracy of detecting R–R intervals, a continuous wavelet transform was implemented. The HRV metrics calculated for each participant, including SD1, SD2, SDNN, RMSSD, and SDSD, were evaluated under two different conditions: "Static subjects" and "Static subjects with makeup". The results showed an average absolute error of 3.53 ms when compared to the ECG chest band device. In addition, the proposed method was compared with the ICA and CHROM methods, and as a result, the proposed method showed better performance in calculating HRV features. Deep learning techniques have become increasingly popular for enhancing the accuracy of HRV analysis. Song et al. [18] introduced the PulseGAN method, which incorporates CHROM and conditional generative adversarial networks (GAN), to estimate HRV from the face. Kuang et al. [19] proposed ESA-rPPGNet, employing 3D depth-wise separable convolution to enhance network performance for HRV analysis. Some researchers used thermal images to extract features to detect stress from the face. Mohd et al. [20] found a correlation between blood flow and temperature changes in facial expression during stress in thermal images. In the research, they used thermal infrared and visible cameras, and the proposed method showed 88.6% accuracy. To increase accuracy, Gioia et al. [21] combined thermal imaging with physiological signals, like cardiac, electrodermal, and respiratory activity, to detect acute stress. All signals were recorded from 25 participants. For classifications, they implemented a support vector machine model. Only using a thermal image system achieved 86.84% of accuracy, and combining it with the physiological features system achieved 97.37% accuracy. Zhang et al. [22] detected stress using a combination of ECG, voice, and facial expressions using deep learning. ECG signal is acquired by three-electrode leads using a Biopac MP160 device; for facial expressions and voice recording, they used a Sony video camera, FDR-AX700. The proposed system showed an accuracy of 0.74 for ECG, 0.83 for voice, and 0.79 for facial expressions. Combining all of them, they can achieve 85.1% accuracy in detecting acute stress. Mitsuhashi et al. [23] combined two methods: the hemoglobin, melanin, and shading (HMS) method and the Spatial Subspace Rotation (2SR) method. A total of 78 videos were collected from 7 subjects. An ECG device was used as ground truth for pulse waves. The participants' stress levels were assessed through responses to the State-Trait Anxiety Inventory (STAI) questionnaire. The K-nearest neighbor method was used for stress classification. The combination of HMS and 2SR systems achieved over 90% accuracy in the relaxation state; in the stress states system, it achieved 80% accuracy.

In this paper, we will evaluate various machine learning techniques to identify the most effective predictive model that can accurately determine stress levels using only HR and HRV features from different datasets. Furthermore, the models with the highest predictive accuracy will be used to classify stress based on HR and HRV features obtained from the face using a camera. Estimation of HRV from the face consists of several main steps: first, we detected the subject's face, and then we employed the plane-orthogonal-to-skin (POS) [24] method. This technique enhances the accuracy of detecting physiological signals from the face. Following this, we applied the discrete wavelet transform (DWT) technique to remove noise from the signal. This step allowed a more precise calculation of HRV from the face. An overview of our proposed method for stress detection from the face is shown in Figure 1. Because stress detection is a complex process and depends on many factors, in our study, we selected three different publicly available datasets, each offering unique insights into stress indicators and responses. These datasets encompass a variety of scenarios, including work-related stress and cognitive tasks. To effectively analyze these datasets and extract meaningful insights into stress indicators, we employed a range of machine learning techniques. These techniques were selected for their particular strengths in pattern recognition and predictive modeling. The first dataset is the SWELL dataset [25]. Many researchers have achieved good prediction accuracy in the SWELL dataset. Sharma et al. [26] applied various machine learning methods, and a two-class neural network model achieved an accuracy of 98% in the SWELL dataset. Another study, conducted by Koldijk et al. [27], achieved 90% accuracy using support vector machines

(SVM). Albaladejo-González et al. [28] achieved an accuracy of 88.64% using a supervised Multi-layer Perceptron (MLP) model. Ghosh et al. [29], using an image-encoding-based deep neural network, achieved a promising accuracy of 99.39% for the SWELL dataset. The second dataset is the PPG sensor dataset. The author implemented various machine learning classifiers, and the K-nearest neighbor (KNN) algorithm achieved 72% accuracy. Using a genetic algorithm led to a significant increase in accuracy, reaching 81% [30]. In the last ECG and EEG sensor dataset [31], the research study focused on classifying stress levels—low stress, moderate stress, and high stress—by analyzing ECG and electroencephalogram (EEG) data. When examining the dataset, the authors found notable differences between genders. Specifically, for females, the accuracy rate of correctly identifying the stress levels was 62.60%, whereas for males, the accuracy increased to 71.57%. To enhance the accuracy of stress classification, the researchers employed stacking techniques. Stacking is a method used in ensemble learning that integrates several classification models to enhance the accuracy of predictions. By using this method, they were able to achieve an overall accuracy of 64.08% across both genders.

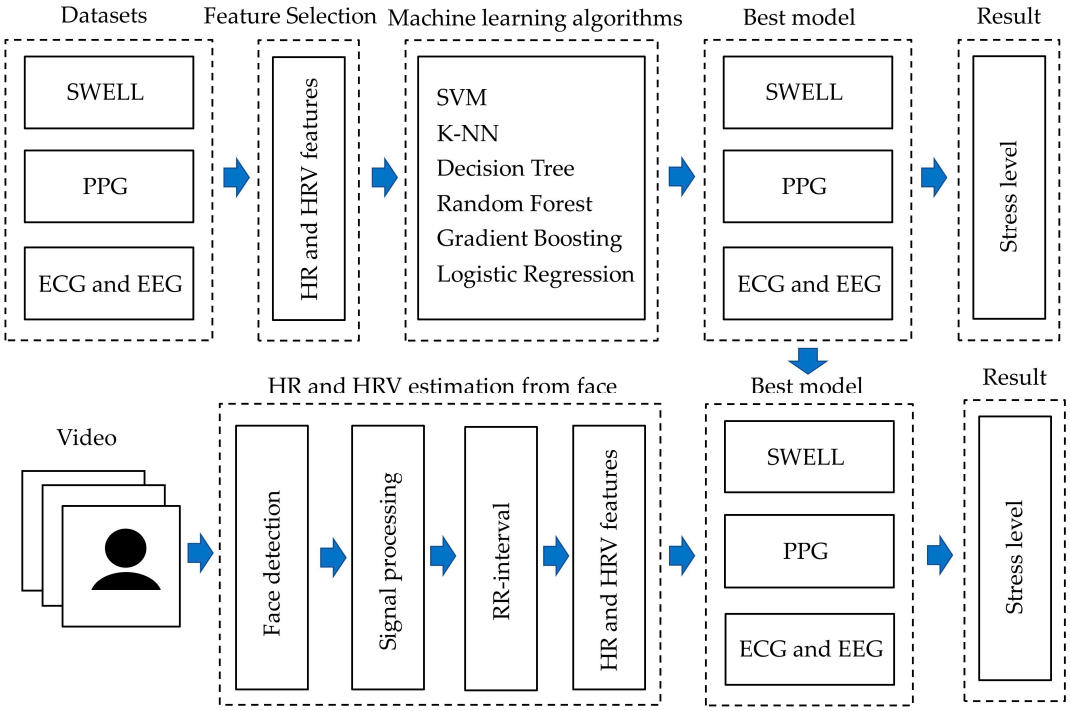


**Figure 1.** An overview of our proposed method.

## 2. Materials and Methods

### *2.1. Datasets*

The SWELL dataset was collected from 17 male and 8 female students, a total of 25 students, in different working conditions. All participants worked under all 3 conditions. In the "Neutral" condition, participants were given the freedom to complete the assigned tasks at their own pace, without any imposed time limits or interruptions. Under the "Time pressure" stressor, participants faced a more challenging scenario. They were required to complete all tasks in only two-thirds of the time it took them to finish under the neutral condition. The "Interruptions" stressor introduced a different type of challenge. Throughout the task, participants received a total of eight emails, deliberately timed to disrupt their workflow and concentration. Each condition in the experiment takes approximately one hour. Before each condition, participants relaxed for 8 min. The dataset contains information regarding computer logs, body postures, facial expressions, ECG signals, and EDA recorded during the experiment.

The real-time diagnosis of mental stress using the photoplethysmography (PPG) sensor dataset was collected from 15 male and 12 female students, for a total of 27 students. The dataset consists of PPG signals collected from the ear lobes of students. A color and word Stroop test was used to determine stressed or normal conditions.

For the ECG and EEG stress features for ECG- and EEG-based detection and multilevel classification of stress using machine learning for specified genders, a preliminary study dataset was collected from 19 male and 21 female students, for a total of 40 students, in different working conditions. The dataset consists of ECG and EEG signals. The mental arithmetic task was used to determine no stress, low stress, and high stress. Each condition was recorded over 5 min. Following the completion of each task, participants were given a 2 min break.

### *2.2. Feature Selection*

In our research, we used HR and HRV features to detect stress levels. HRV features and definitions are shown in Table 1.

**Table 1.** HRV features and definitions.

| Feature | Unit | Definition |
|---|---|---|
| MEAN_RR | ms | Mean of RR intervals |
| MEDIAN_RR | ms | Median of RR intervals |
| SDRR | ms | The standard deviation of RR intervals |
| RMSSD | ms | Root mean square of successive RR interval differences |
| SDSD | ms | The standard deviation of successive RR interval differences |
| pNN50 | % | Percentage of successive RR intervals that differ by more than 50 ms |
| pNN25 | % | Percentage of successive RR intervals that differ by more than 50 ms |
| SD1 | ms | Poincaré plot standard deviation perpendicular the line of identity |
| SD2 | ms | Poincaré plot standard deviation along the line of identity |
| VLF | $ms^2$ | Very-low-frequency band (0.0033–0.04 Hz) |
| LF | Hz | Low-frequency band (0.04–0.15 Hz) |
| HF | Hz | High-frequency band (0.15–0.4 Hz) |
| LF/HF | % | The ratio of LF-to-HF frequency |

### *2.3. Extracting Heart Rate (HR) and Heart Rate Variability (HRV) from the Face*

Initially, the process begins by identifying and selecting a whole face region; to achieve accurate face detection we used Mediapipe (0.9.0.1) [32]. After the face was detected, we converted video frames to YCrCb color spaces. These color spaces offer distinct advantages for processing and analyzing visual information related to human skin, particularly in extracting subtle color variations that correspond to physiological changes. The combination of these color spaces helps to mitigate issues related to variable lighting conditions and skin tone diversity, enhancing the accuracy of skin detection. Then, the average RGB values are computed from the selected ROI. After that, we used the plane-orthogonal-to-skin (POS) method. This method is designed for extracting a robust signal for blood volume pulse measurement from video by projecting the RGB color space signals in a plane orthogonal to the skin tone direction. The POS method transforms the RGB signals into a new space where the components are more representative of the blood volume pulse (BVP) signal, while reducing noise, including motion artifacts and changes in ambient lighting. Then, detrending and filtering are applied to remove noise and emphasize the frequency band of interest. Detrending and filtering are essential for removing low-frequency trends and high-frequency noise, focusing on the frequency band corresponding to typical heart rates. In our study, we chose the discrete wavelet transform (DWT) Daubechies 4 with decompo-

sition level 5 for signal denoising because of its high effectiveness in removing noise from signals [33]. The DWT has demonstrated notable success in not only cleaning noise from signals but also in reducing motion artifacts, particularly in photoplethysmography (PPG) and rPPG signals [34]. Motion artifacts represent a significant challenge in rPPG signal processing, as they can significantly distort signals. The continuous wavelet transform (CWT) for a signal $f(t)$ is defined as

$$\mathrm{CWT}_{(a,b)} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi^*(a,b) dt \tag{1}$$

where

$$\Psi^*(a,b) = \frac{1}{\sqrt{a}} \Psi \left( \frac{t-b}{a} \right) \tag{2}$$

$a$ is a scale factor and $b$ is a translation factor; $\Psi$ is the mother wavelet; and the DWT can be defined as

$$\mathrm{DWT}_{t(j,k)} = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t) \Psi^*(j,k) dt \tag{3}$$

$$\Psi^*(j,k) = \frac{1}{\sqrt{2^j}} \Psi \left( \frac{t-k2^j}{2^j} \right) \tag{4}$$

where the $2^j$ scale and the $k2^j$ shift parameters, $j, k \in Z$.

DWT denoising consists of three main steps: decomposition, thresholding, and signal reconstruction, see Figure 2. DWT decomposes a signal into approximation coefficients (cAn) and detailed coefficients (cDn), Figure 3. Thresholding effectively smooths the signal by reducing the noise that lies near the threshold. In the final step, the inverse discrete wavelet transform is used to reconstruct the signal.

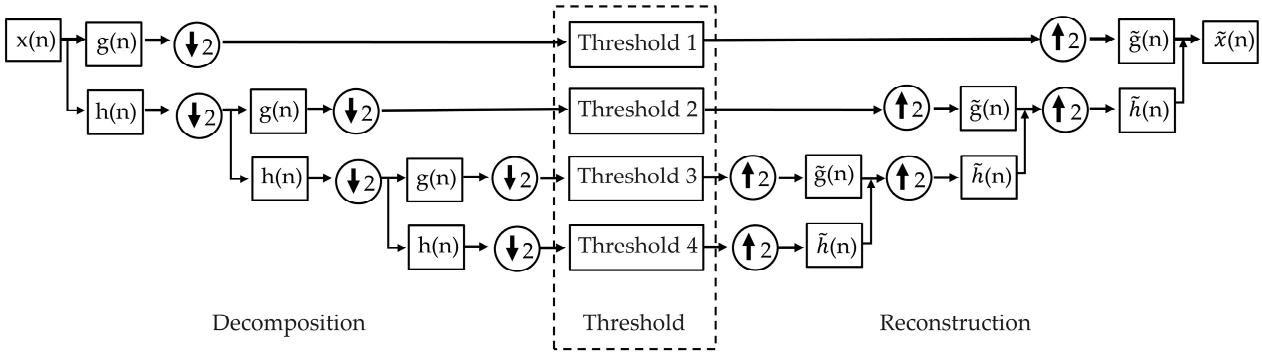


**Figure 2.** The rPPG denoising process decomposition level 3; h(n) is the low-pass filter and g(n) is the high-pass filter; 2 is downsampling and upsampling factor.

DWT stands out for its ability to deconstruct signals into various frequency components, allowing for the precise identification and removal of noise while preserving the main features of the original signal. This characteristic makes DWT an invaluable tool in the processing of physiological signals, where maintaining the main features of the data is paramount.

This approach allows for a cleaner extraction and analysis of physiological data, essential for accurate HRV measurements. This filtering technique preserves the signal's essential features while reducing noise. Normalization is used to scale the smoothed signal to a standard range (Figure 4). This step prepares the data for peak detection by ensuring that the signal amplitude does not bias the detection process. Peak detection in the processed signal is used to identify heartbeats. Finally, the inter-beat interval (IBI) is calculated by taking the difference in time between successive peaks identified in the normalized signal (Figure 5). After successfully identifying the IBI interval, we calculated all the HRV features shown in Table 1.

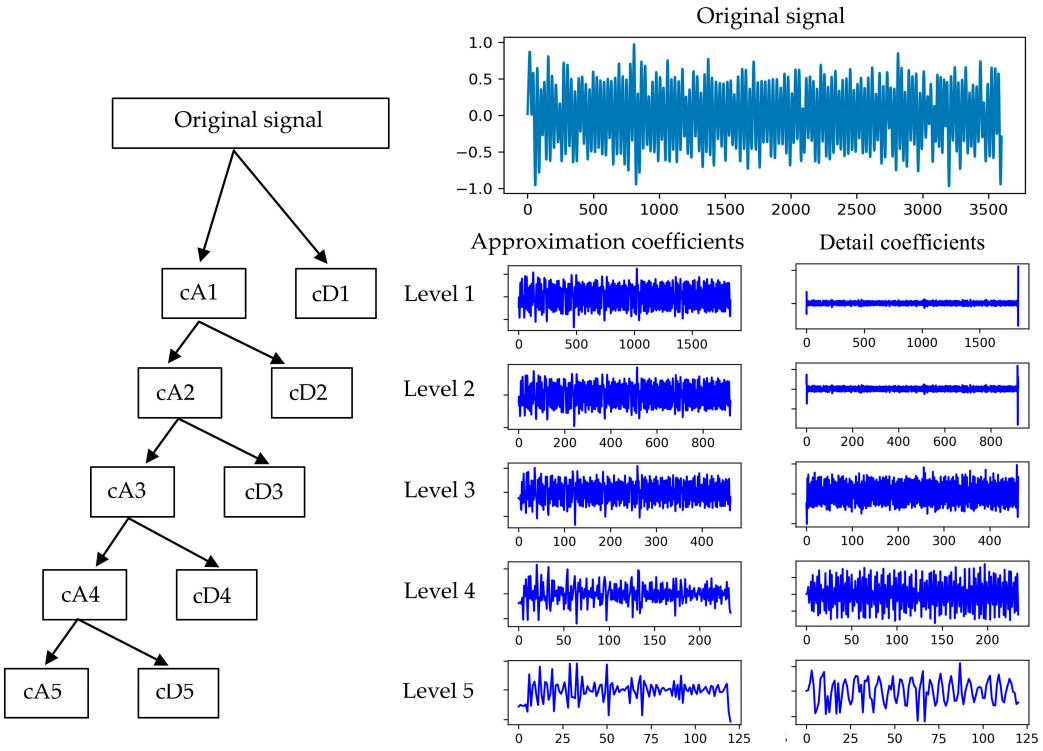


**Figure 3.** Multi-level wavelet decomposition diagram.

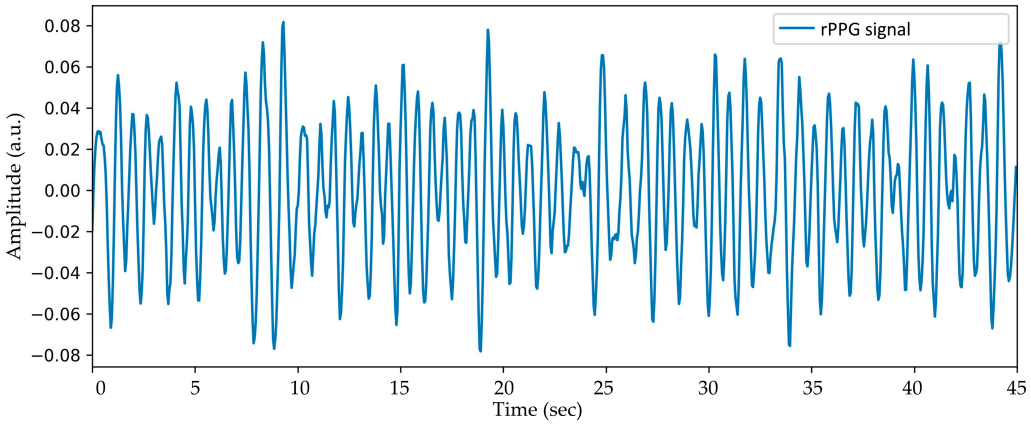


**Figure 4.** Normalized signal.

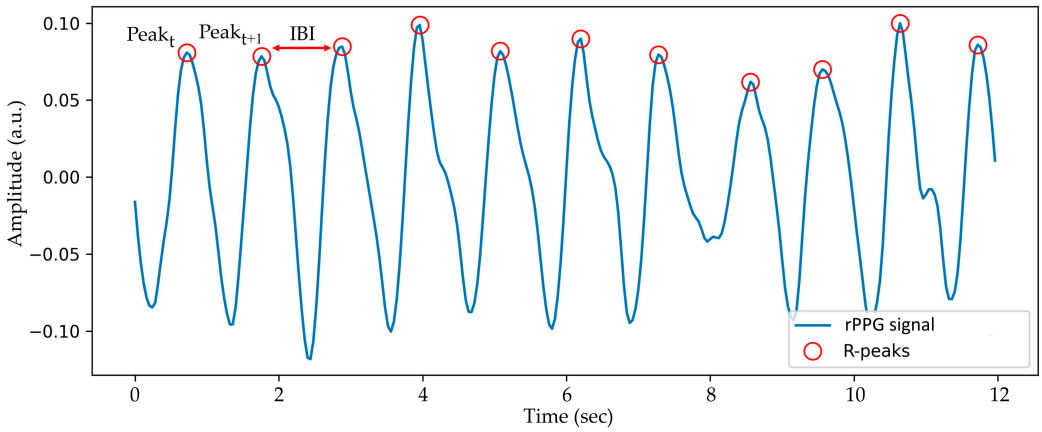


**Figure 5.** Peak detection and IBI calculation from rPPG signal obtained from face.

To validate the accuracy of the heart rate (HR) and heart rate variability (HRV) obtained from the camera-based system, we tested our algorithm on the publicly available PURE dataset [35] and a custom dataset. The pulse rate detection dataset of the PURE dataset consists of 10 subjects, 6 different head motion setups (steady, talking, slow translation, fast translation, small rotation, and medium rotation of head), with a total of 60 videos. The videos were recorded with an eco274CVGE camera by SVS-Vistek GmbH, a resolution of 640 × 480 pixels at 30 frames per second, and a finger pulse oximeter (pulox CMS50E) used to simultaneously record subjects' ground truth data.

For the custom dataset, a fingertip pulse oximeter (YK-82C, Yonker, Xuzhou, China) was used as the ground truth device for HR. The ground truth HRV signal was collected using the Polar H10 (Polar Electro, Kempele, Finland) chest strap attached to the subject's chest. The accuracy of chest straps is often comparable to that of clinical-grade equipment, making them an excellent choice for reliable HRV data without the need to visit a clinic [36]. Each subject was seated approximately 60 cm in front of a camera, and time-synchronized facial video was recorded with a C920 HD PRO (Logitech, Suzhou, China) webcam for 2 min, at 30 frames per second (fps) with a resolution of 1920 × 1080 pixels, Figure 6a. A total of 20 videos were collected from five participants. Before the experiment, the participants were asked to rest for 5 min. Participants were instructed to avoid caffeine, exercise, and other factors that could significantly affect their heart rate or stress levels for at least 2 h before the experiment. Two hours before the public speaking, the first measurements of each participant's HR and HRV were taken, for the physiological state of each participant in a rested condition. For stress conditions, participants were asked to perform a public speaking task. During the public speaking task, each participant's face was recorded using cameras, as shown in Figure 6b.

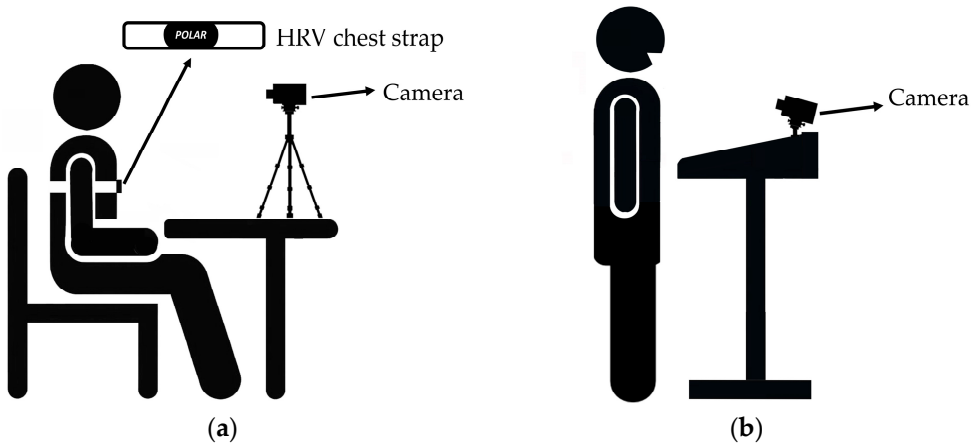


**Figure 6.** Overview of the experimental setup: (**a**) rested condition and (**b**) public speaking.

## 3. Results

In our study, we used only the HR and HRV features from the publicly available dataset. and applied various machine learning techniques. These models included a decision tree, logistic regression, random forest, K-nearest neighbor, gradient-boosting classifier, and support vector classifier with a linear kernel. To provide a rigorous assessment of the model's performance and to enhance the generalizability of our findings, we employed 10-fold cross-validation. For each model, we conducted a series of experiments to fine-tune its parameters, aiming to achieve the best possible balance between sensitivity and specificity.

In the SWELL dataset, random forest achieved an impressive 99% predictive accuracy in stress level prediction by using only HR and HRV features. To achieve this, the dataset was divided into a training set (70%) and a testing set (30%). We normalized the training and testing datasets by scaling feature values to a range between 0 and 1. This normalization helps in speeding up the learning process and improves the performance of many

machine learning algorithms by eliminating the bias that can occur due to the variance in measurement scales. In this study, we employed the minimal-redundancy-maximal-relevance (mRMR) feature selection technique. It helps in identifying a compact set of features that contribute most significantly to predicting the outcome. We selected eight features using the mRMR approach. The hyperparameter configuration of the random forest classifier is as follows: the classifier was configured with 100 trees, a maximum depth of 15, and leaves containing at least one sample. This high level of accuracy demonstrates the capability to predict stress effectively using only HR and HRV metrics. Table 2 demonstrates the performance of the random forest model in classifying stress levels. Table 3 demonstrates the performance comparison of different models, and Table 4 demonstrates a comparison of the proposed method with existing methods.

**Table 2.** Performance of the random forest model for three classifications levels.

| Classifications Level | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|
| No stress | 0.99 | 0.99 | 0.99 | 0.99 |
| Interruption | 0.99 | 0.99 | 0.99 | 0.99 |
| Time pressure | 0.99 | 0.98 | 0.98 | 0.99 |

**Table 3.** Performance comparison of different models (SWELL).

| Dataset | Accuracy |
|---|---|
| Decision Tree | 97% |
| Logistic Regression | 64% |
| K-Nearest Neighbor | 74% |
| Gradient-Boosting Classifier | 86% |
| Support Vector Classifier | 63% |
| Random Forest | 99% |

**Table 4.** Comparison of the proposed method with existing methods on SWELL dataset.

| Dataset | Accuracy |
|---|---|
| Sharma et al. [26] | 98% |
| Koldijk et al. [27] | 90% |
| Albaladejo-González et al. [28] | 88.64% |
| Ours | 99% |

In the PPG sensor dataset among the tested models, the logistic regression technique outperformed other methodologies in terms of predictive accuracy, achieving a good accuracy rate of 84.2%. In the study, data preparation and analysis involved several critical stages. In our study, the initial stages of data analysis involved cleaning, preprocessing, and filtering of the raw data. Following filtering, we extracted both time-domain and frequency-domain HRV features. The dataset was then split into training and testing sets in a 70/30 ratio. After splitting, we normalized the data. The logistic regression model was configured with a maximum iteration parameter of 2000. Other parameters of the logistic regression model were left at their default settings. Table 5 demonstrates the performance of the logistic regression model in classifying stress levels. Table 6 demonstrates a performance comparison of different models, and Table 7 demonstrates a comparison of the proposed method with existing methods.

**Table 5.** Performance of the logistic regression model for two classification levels.

| Classification Level | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|
| Normal | 0.86 | 0.80 | 0.80 | 0.84 |
| Stressed | 0.83 | 0.89 | 0.84 | 0.84 |

**Table 6.** Performance comparison of different models (PPG sensor dataset).

| Dataset | Accuracy |
|---|---|
| Decision Tree | 62% |
| Random Forest | 73% |
| K-Nearest Neighbor | 71% |
| Gradient-Boosting Classifier | 78% |
| Support Vector Classifier | 75% |
| Logistic Regression | 84% |

**Table 7.** Comparison of the proposed method with existing methods on the PPG sensor dataset.

| Dataset | Accuracy |
|---|---|
| Talha et al. [30] | 81% |
| Ours | 84% |

For ECG and EEG datasets, the datasets were split into training and testing sets in an 80/20 ratio. We employed a soft voting ensemble technique, combining the predictions of diverse classifiers to enhance the robustness and accuracy of stress prediction. A variety of base learners were initialized, including logistic regression, decision tree, random forest, and K-nearest neighbors. All classifiers used default parameters as a starting point, providing a balanced approach between performance and computational efficiency. These classifiers were integrated into a soft voting ensemble, wherein each classifier's probabilistic predictions were aggregated to yield a final verdict. This approach capitalizes on the strength of each base learner and compensates for their weaknesses. This approach significantly enhanced our predictive accuracy, leading to a notable improvement, with the ensemble model achieving an accuracy rate of 67%. Table 8 demonstrates the performance of the ensemble model in classifying stress levels. Table 9 demonstrates a performance comparison of different models, and Table 10 demonstrates a comparison of the proposed method with existing methods.

**Table 8.** Performance of the ensemble model for three classification levels.

| Classifications Level | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|
| Normal | 0.67 | 0.67 | 0.69 | 0.67 |
| Low stress | 0.67 | 0.67 | 0.66 | 0.66 |
| High stress | 0.66 | 0.68 | 0.65 | 0.68 |

To validate the proposed method for estimating HR and HRV from facial data, we calculated the mean absolute error (MAE) for both the standard deviation of normal-to-normal intervals (SDNN) and heart rate. We conducted experiments using the PURE datasets. On the PURE dataset, our HR and HRV estimation models achieved an MAE of 9.32 ms for SDNN and an MAE of 1.75 bpm for heart rate. The results of these experiments are presented in Figures 7 and 8 and Table 11.

**Table 9.** Performance comparison of different models (ECG and EEG dataset).

| Dataset | Accuracy |
|---|---|
| Decision Tree | 63% |
| Random Forest | 50% |
| K-Nearest Neighbor | 42% |
| Gradient-Boosting Classifier | 58% |
| Support Vector Classifier | 48% |
| Logistic Regression | 62% |
| Ensemble Model | 67% |

**Table 10.** Comparison of the proposed method with existing methods on ECG and EEG datasets.

| Dataset | Accuracy |
|---|---|
| Hemakom et al. [31] | 64.08% |
| Ours | 67% |

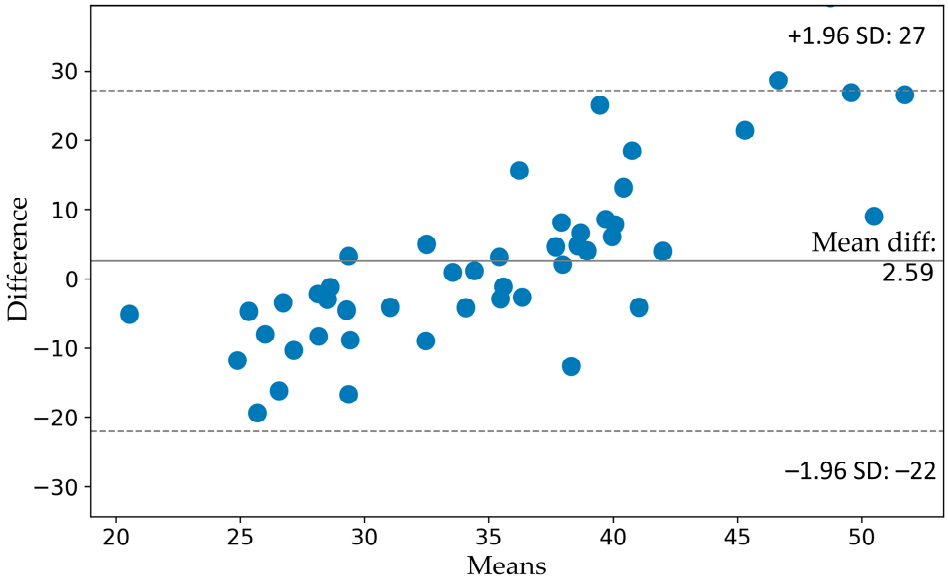


**Figure 7.** Bland–Altman plots between the predicted SDNN and the ground truth SDNN on PURE dataset.

**Table 11.** A comparison of the performance of various methods on the PURE dataset.

| Method | $\mathbf{SDNN_{MAE}}$ (ms) |
|---|---|
| CHROM | 89.30 |
| PulseGAN | 49.39 |
| FaceRPPG | 18 |
| PhysNet | 14.22 |
| rPPGGAN | 12.56 |
| ESA-rPPGNet | 11.75 |
| Ours | 9.32 |

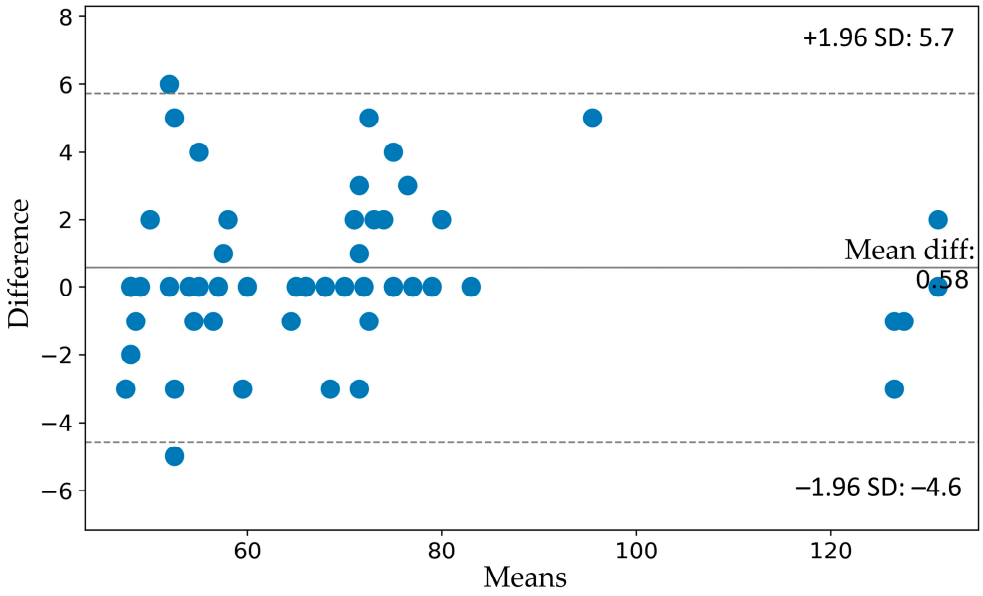


**Figure 8.** Bland–Altman plots between the predicted HR and the ground truth HR on PURE dataset.

As shown in Table 11, our proposed method performed better than traditional CHROM [18], FaceRPPG [19] methods, and deep learning-based PulseGAN [18], PhysNet, [19], rPPGGAN [37], ESA-rPPGNet [19] methods. In the custom dataset, our proposed system showed an accuracy of 95% for HR and 82% for overall HRV features Table 12.

**Table 12.** The accuracy of the proposed system, HRV calculated from the face compared with ground truth Polar H10 chest strap data.

| MEAN_RR | SDRR | RMSSD | SDSD | pNN50 | pNN25 | SD1 | SD2 | VLF | LF | HF | LF/HF |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 84% | 83% | 88% | 82% | 81% | 79% | 80% | 83% | 76% | 79% | 81% | 82% |

To evaluate the models under real-world conditions, we measured the participants' stress levels both in a normal state and during public speaking. Public speaking is widely recognized as an effective mental stressor, making it an ideal method for inducing stress [9]. During the public speaking task, each participant's face was recorded using cameras. An experiment was conducted with 5 participants. We evaluated the performance of models trained on different datasets: a SWELL dataset, a PPG dataset, and an ECG and EEG dataset. Each dataset offers physiological signals and their correlation with stress.

We evaluated participants' stress levels using the best models before public speaking activities. The results showed that, at this stage, none of the participants demonstrated any signs of stress. For the random forest model trained on the SWELL dataset, an impressive detection rate was observed, with all 5 participants showing signs of stress (interruption, time pressure) during the public speaking task. This suggests that the SWELL dataset, which includes a variety of physiological responses to stressors, is particularly effective for training models to recognize stress during high-pressure tasks like public speaking. On the other hand, the logistic regression model trained on the second PPG datasets identified stress in 3 out of the 5 participants. The detection rate of 60% indicates a significant sensitivity to stress. For the ensemble model trained on the ECG dataset, no stress was detected in any of the 5 participants. This result could suggest several possibilities, such as the model achieved an accuracy rate of 67% during its training phase. It also suggests that the model's detection capabilities are somewhat limited, potentially leading to an inability to accurately recognize stress under certain conditions. Secondly, the dataset size used for training this model was small, which significantly impacted its performance. Consequently,

the small dataset size, combined with the moderate accuracy, contributed to the model's failure to detect stress in the participants.

## 4. Discussion

Stress is increasingly being recognized as a significant issue with serious consequences for health. In this research, we proposed a contactless method for stress detection using a standard camera. We used only HR and HRV to determine stress. Using only HRV, we achieved a high result, especially in the SWELL dataset. This suggests that HRV is a powerful indicator of stress. The high performance of a model trained on the SWELL dataset, in comparison to models trained on other datasets, is because of the dataset's extensive size and the comprehensive duration over which the signals were recorded. The SWELL dataset comprises 410,322 records. A large number of examples provide more information about stress characteristics, increasing the model's ability to detect stress in different situations. This large dataset size is critical to training robust machine learning models because it allows the identification of subtle patterns in the data that smaller datasets may miss. PPG and ECG datasets could not capture the full spectrum of stress responses, potentially limiting the model's learning scope.

In summary, the superior performance of the model trained on the SWELL dataset is a direct consequence of the dataset's large size and the comprehensive nature of the physiological signals it contains, recorded over a significant duration. This analysis highlights the importance of dataset selection in developing highly effective stress detection models and suggests that the inclusion of diverse and extensive data can significantly enhance model performance. While the current study presents a promising contactless method for stress detection via HRV analysis using a simple camera, there are some limitations that affect the accuracy of the system. One of the primary limitations identified in our proposed system is the sensitivity to environmental factors, specifically the influence of lighting conditions during the process of assessing HRV using a camera. This sensitivity presents a significant challenge, as fluctuations in light intensity and direction can affect the camera's ability to detect the changes in skin coloration associated with heartbeats. Such conditions can lead to inaccuracies in HRV readings, potentially impacting the system's overall effectiveness. Another limitation is that sharp movements and dark skin tones can significantly influence the system's accuracy. While the system is less sensitive to minor, subtle motions, sudden or sharp movements can still disrupt the signal. Moreover, dark skin tones can also pose challenges due to their lower reflectivity and higher absorption of light. As a result, both sharp movements and dark skin tones can lead to decreased accuracy in the measurements. Future research will focus on addressing this limitation and exploring stress responses across various scenarios with more participants, extending the investigation to include other vital sign indicators.

## 5. Conclusions

In conclusion, in this study, we have proposed a camera-based stress detection system based on HR and HRV, calculated from the face. HR and HRV estimations from the face were validated on the PURE public dataset and the private custom dataset. Compared with other methods, the proposed method showed better results in estimating SDNN. In addition, we examined three distinct datasets using various machine learning techniques. This approach allowed us to identify the most effective methods from each dataset for stress detection. Following this evaluation, we further tested the best-performing techniques under real-life conditions to validate their effectiveness and reliability in accurately identifying stress. The positive results from these models demonstrated the efficacy of our approach to calculating HRV from facial video. Our system provides a continuous, cost-effective solution for stress detection using an ordinary camera. Continuously monitoring stress levels is particularly important for individuals suffering from high blood pressure and heart disease. Regular monitoring allows for the timely identification of stress, and they can implement stress-reduction techniques to enhance overall health.

**Author Contributions:** Conceptualization, J.-H.L., D.L. and M.K.; methodology, J.-H.L.; software, M.K.; validation, J.-H.L. and M.K.; formal analysis, M.K.; investigation, J.-H.L. and C.-H.K.; resources, J.-H.L., C.-H.K. and M.K.; data curation, J.-H.L., D.L., C.-H.K. and M.K.; writing—original draft preparation, J.-H.L., C.-H.K. and M.K.; writing—review and editing, J.-H.L., C.-H.K. and M.K.; visualization, J.-H.L., D.L., M.K. and C.-H.K.; supervision, J.-H.L., C.-H.K. and M.K.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2023(DBSD1-06), Basic Research Program through the National Research Foundation of Korea (NRF-2022R1I1A307278), Korea Health Technology R&D Project through the Korea Health Industry Development Institute(KHIDI)(HI21C0977, RS-2021-KH118978, RS-2024-00433896) and Korea Medical Device Development Fund grant funded by the Korea government(RS-2022-00166898).

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Vancheri, F.; Longo, G.; Vancheri, E.; Henein, M.Y. Mental Stress and Cardiovascular Health-Part I. *J. Clin. Med.* **2022**, *11*, 3353. [CrossRef] [PubMed]
2. Kulkarni, S.; O'Farrell, I.; Erasi, M.; Kochar, M.S. Stress and hypertension. *WMJ* **1998**, *11*, 34.
3. Guiraud, V.; Amor, M.B.; Mas, J.L.; Touzé, E. Triggers of ischemic stroke: A systematic review. *Stroke* **2010**, *41*, 2669–2677. [CrossRef] [PubMed]
4. Sun, Y.; Yu, X. An innovative nonintrusive driver assistance system for vital signal monitoring. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1932–1939. [CrossRef]
5. Castaldo, R.; Melillo, P.; Bracale, U.; Caserta, M.; Triassi, M.; Pecchia, L. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomed. Signal Process. Control.* **2015**, *18*, 370–377. [CrossRef]
6. Posada-Quintero, H.F.; Florian, J.P.; Orjuela-Cañón, A.D.; Chon, K.H. Electrodermal Activity Is Sensitive to Cognitive Stress under Water. *Front. Physiol.* **2018**, *8*, 1128. [CrossRef]
7. Tzevelekakis, K.; Stefanidi, Z.; Margetis, G. Real-Time Stress Level Feedback from Raw Ecg Signals for Personalised, Context-Aware Applications Using Lightweight Convolutional Neural Network Architectures. *Sensors* **2021**, *21*, 7802. [CrossRef] [PubMed]
8. Keshan, N.; Parimi, P.V.; Bichindaritz, I. Machine learning for stress detection from ECG signals in automobile drivers. In Proceedings of the IEEE International Conference on Big Data, Santa Clara, CA, USA, 29 October–1 November 2015; pp. 2661–2669. [CrossRef]
9. Lee, S.; Hwang, H.B.; Park, S.; Kim, S.; Ha, J.H.; Jang, Y.; Hwang, S.; Park, H.-K.; Lee, J.; Kim, I.Y. Mental Stress Assessment Using Ultra Short Term HRV Analysis Based on Non-Linear Method. *Biosensors* **2022**, *12*, 465. [CrossRef] [PubMed]
10. Zubair, M.; Yoon, C. Multilevel Mental Stress Detection Using Ultra-Short Pulse Rate Variability Series. *Biomed. Signal Process. Control.* **2020**, *57*, 101736. [CrossRef]
11. Li, Z.; Snieder, H.; Su, S.; Ding, X.; Thayer, J.F.; Treiber, F.A.; Wang, X. A longitudinal study in youth of heart rate variability at rest and in response to stress. *Int. J. Psychophysiol.* **2009**, *73*, 212–217. [CrossRef]
12. Tharion, E.; Parthasarathy, S.; Neelakantan, N. Short-term heart rate variability measures in students during examinations. *Natl. Med. J. India* **2009**, *22*, 63–66. [PubMed]
13. Taelman, J.; Vandeput, S.; Vlemincx, E.; Spaepen, A.; Van Huffel, S. Instantaneous changes in heart rate regulation due to mental load in simulated office work. *Eur. J. Appl. Physiol.* **2011**, *111*, 1497–1505. [CrossRef]
14. Visnovcova, Z.; Mestanik, M.; Javorka, M.; Mokra, D.; Gala, M.; Jurko, A.; Calkovska, A.; Tonhajzerova, I. Complexity and time asymmetry of heart rate variability are altered in acute mental stress. *Physiol. Meas.* **2014**, *35*, 1319–1334. [CrossRef] [PubMed]
15. Traina, C.A.; Galullo, M.; Russo, F.G. Effects of anxiety due to mental stress on heart rate variability in healthy subjects. *Minerva Psichiatr.* **2011**, *52*, 227–231.
16. Luo, J.; Zhang, G.; Su, Y.; Lu, Y.; Pang, Y.; Wang, Y.; Wang, H.; Cui, K.; Jiang, Y.; Zhong, L.; et al. Quantitative analysis of heart rate variability parameter and mental stress index. *Front. Cardiovasc. Med.* **2022**, *9*, 930745. [CrossRef] [PubMed]
17. Huang, R.-Y.; Dung, L.-R. Measurement of heart rate variability using off-the-shelf smart phones. *Biomed. Eng. Online* **2016**, *15*, 11. [CrossRef]
18. Song, R.; Chen, H.; Cheng, J.; Li, C.; Liu, Y.; Chen, X. PulseGAN: Learning to Generate Realistic Pulse Waveforms in Remote Photoplethysmography. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1373–1384. [CrossRef] [PubMed]
19. Kuang, H.; Lv, F.; Ma, X.; Liu, X. Efficient Spatiotemporal Attention Network for Remote Heart Rate Variability Analysis. *Sensors* **2022**, *22*, 1010. [CrossRef] [PubMed]
20. Mohd, M.N.H.; Kashima, M.; Sato, K.; Watanabe, M. Mental stress recognition based on non-invasive and non-contact measurement from stereo thermal and visible sensors. *Int. J. Affect. Eng.* **2015**, *14*, 9–17. [CrossRef]

21. Gioia, F.; Greco, A.; Callara, A.L.; Scilingo, E.P. Towards a Contactless Stress Classification Using Thermal Imaging. *Sensors* **2022**, *22*, 976. [CrossRef]
22. Zhang, J.; Yin, H.; Zhang, J.; Yang, G.; Qin, J.; He, L. Real-time mental stress detection using multimodality expressions with a deep learning framework. *Front. Neurosci.* **2022**, *16*, 947168. [CrossRef] [PubMed]
23. Mitsuhashi, R.; Iuchi, K.; Goto, T.; Matsubara, A.; Hirayama, T.; Hashizume, H.; Tsumura, N. Video-Based Stress Level Measurement Using Imaging Photoplethysmography. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July 2019; pp. 90–95. [CrossRef]
24. Wang, W.; den Brinker, A.C.; Stuijk, S.; de Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1479–1491. [CrossRef] [PubMed]
25. Koldijk, S.; Sappelli, M.; Verberne, S.; Neerincx, M.A.; Kraaij, W. The SWELL Knowledge Work Dataset for Stress and User Modeling Research. In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI'14), Istanbul, Turkey, 12–16 November 2014; pp. 291–298. [CrossRef]
26. Sharma, R.; Rani, S.; Gupta, D. Stress detection using machine learning classifiers in internet of things environment. *J. Comput. Theor. Nanosci.* **2019**, *16*, 4214–4219. [CrossRef]
27. Koldijk, S.; Neerincx, M.A.; Kraaij, W. Detecting Work Stress in Offices by Combining Unobtrusive Sensors. *IEEE Trans. Affect. Comput.* **2018**, *9*, 227–239. [CrossRef]
28. Albaladejo-González, M.; Ruipérez-Valiente, J.A.; Gómez Mármol, F. Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate. *J. Ambient. Intell. Human Comput.* **2023**, *14*, 11011–11021. [CrossRef]
29. Ghosh, S.; Kim, S.; Ijaz, M.F.; Singh, P.K.; Mahmud, M. Classification of Mental Stress from Wearable Physiological Sensors Using Image-Encoding-Based Deep Neural Network. *Biosensors* **2022**, *12*, 1153. [CrossRef] [PubMed]
30. Anwar, T.; Zakir, S. Machine Learning Based Real-Time Diagnosis of Mental Stress Using Photoplethysmography. *J. Biomim. Biomater. Biomed. Eng.* **2022**, *55*, 154–167. [CrossRef]
31. Apit, H.; Danita, A.; Pasin, I. ECG and EEG Stress Features for: ECG and EEG Based Detection and Multilevel Classification of Stress Using Machine Learning for Specified Genders: A Preliminary Study [Dataset]. 2023, Dryad. Available online: https://datadryad.org/stash/dataset/doi:10.5061/dryad.kd51c5bbf (accessed on 10 January 2024).
32. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.; Yong, M.; Lee, J.; et al. Mediapipe: A framework for building perception pipelines. *arXiv* **2019**, arXiv:1906.08172.
33. Haddadi, R.; Abdelmounim, E.; El Hanine, M.; Belaguid, A. Discrete Wavelet Transform based algorithm for recognition of QRS complexes. In Proceedings of the 2014 International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, 14–16 April 2014; pp. 375–379. [CrossRef]
34. Abdulrahaman, L.Q. Two-Stage Motion Artifact Reduction Algorithm for rPPG Signals Obtained from Facial Video Recordings. *Arab. J. Sci. Eng.* **2024**, *49*, 2925–2933. [CrossRef]
35. Stricker, R.; Müller, S.; Gross, H.M. Non-contact video-based pulse rate measurement on a mobile service robot. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1056–1062.
36. Schaffarczyk, M.; Rogers, B.; Reer, R.; Gronwald, T. Validity of the Polar H10 Sensor for Heart Rate Variability Analysis during Resting State and Incremental Exercise in Recreational Men and Women. *Sensors* **2022**, *22*, 6536. [CrossRef]
37. Kuang, H.; Ao, C.; Ma, X.; Liu, X. Remote photoplethysmography signals enhancement based on generative adversarial networks. In Proceedings of the 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 26–28 May 2023; pp. 792–796. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.