

Non-contact Facial based Vital Sign Estimation using Convolutional Neural Network Approach

Nor Surayahani Suriani, Nur Syahida Shahdan, Nan Md. Sahar, Nik Shahidah Afifi Md. Taujuddin
Dept. of Electronics Engineering, Universiti Tun Hussein Onn Malaysia
Batu Pahat, Johor, Malaysia

Abstract—A rapid heart rate may indicate early diagnosis of heart disease, which could result in abrupt mortality if a heart attack occurs while exercising. A fatal incident is usually precipitated by a heart attack while strenuously exercising. This paper proposed invasive health monitoring through remote photoplethysmography (rPPG) analysis captured by RGB video camera to measure a wide range of biological data. A non-contact facial-based vital signs prediction can facilitate checking pulse rate and respiration rate regularly. Several studies have been conducted on evaluating rPPG signals under a variety of static conditions and little head movement, including different skin tones, angles of the camera, and distance from the camera. A study of heart rate (HR) and breathing rate (BR) data from facial videos for fitness applications were presented in this paper. Most studies still do not have a way to measure vital sign estimation especially for physical activity application from facial videos. The face detector was applied based on three regions of interest on facial landmarks for vital sign estimation. Then, the rPPG method with convolutional neural network (CNN) is presented to construct a spatio-temporal mapping of essential characteristics and estimate the vital sign from a sequence of facial images of people after doing various types of exercises. This will allow people to keep track of their health while exercising and creating a tailored training program based on their physiological preferences. The absolute error (AE) between the estimated HR and the reference HR from all experiments is 2.16 ± 2.2 beats/min. While the AE for the estimated BR from the references BR are 1.53 ± 2.3 beats/min.

Keywords—rPPG; remote heart rate estimation; respiration rate; fitness

I. INTRODUCTION

HR and BR are key health indicators for monitoring heart and lung function, especially during this critical COVID-19 pandemic period. Monitoring post-COVID-19 patients' physical activity responses helps in the recovery process. To monitor vital indicators including respiration rate, oxygen saturation, and heart rate, contactless measurement is critical. Infections, fever, asthma, and breathing issues are all signs of a higher HR. Pneumonia or lung illness might create an unstable BR. BR can be influenced by a variety of circumstances, including exercise, emotions, and injuries. Shortness of breath and cough was noted by nearly 10% and 25% of COVID-19 patients, respectively [1].

Our research goal is to assess the efficacy of our prototype of health monitoring status, which will include a built-in camera for non-contact vital sign estimation. The aim of this study presented in this article shows the potential of

estimating HR and BR from facial video streams. We also intended to validate the performance of our prototype so that it could be integrated with the CNN model as a comprehensive non-invasive system to aid in the creation of our healthcare system. The proposed work demonstrates the efficacy of a CNN-based model for post-exercise datasets. The dataset was created to expand the application of our prototype health monitoring system.

Remote photoplethysmography (rPPG) approaches use a webcam or a mobile camera to determine an individual's heartbeat from pixel variation in human skin surface induced by cardiac activity. The principle that blood absorbs more light than other surrounding tissue is used to investigate variations in blood volume transmission on the skin surface using an optics-based rPPG approach. In general, the rPPG procedure entails detecting and tracking the individuals' skin colour changes. Heart rate, heart rate variability, and respiration rate were evaluated using the tracking signals.

Several studies have reported that the potential of rPPG in heart rate estimation is quite promising. Reflected light and movement interferences, however, continue to cause significant challenges. Furthermore, rather than BR, rPPG signals are frequently applied to measure HR. This is because rPPG's frequency properties may not be reliable in estimating respiratory rate. Traditional rPPG approaches have some drawbacks because they rely on assumptions like simplified noise reduction and a skin optical reflection model. Furthermore, remote rPPG approaches will produce unstable results in real-world circumstances where patients are in motion. Deep learning breakthroughs have considerably enhanced the performance of rPPG approaches [2].

This paper describes how to use spatial-temporal facial features mapping to train the CNN model and estimate the rPPG matrix vector on the network's final layers. The single vector represents the rPPG signal, which will be processed further to estimate the HR and BR. Compared to previous work [11-13] and [16], our method addresses the efficacy of vital sign estimation from facial video for fitness or involving physical activity. Our proposed method also addresses the challenges to improve the system robustness. As a result, the following are the paper's contributions:

1) Limit the facial regions to only the forehead, cheeks, and nose to form spatial-temporal features for the CNN model to increase the reliable information for blood variation on face. Our approach improves the signal-to-noise ratio (SNR) for a better signal quality assessment of physiological signals.

2) Calculate HR and BR from the estimated rPPG signal, which is the CNN output. Our approach is in line with deep learning research community where many CNN frameworks have been successfully developed for detection and estimation. Further signal processing algorithms are proposed to estimate the HR and BR from the estimated rPPG signals.

3) For post-exercises with more variable heart rates, we use the rPPG signal to estimate HR and BR using CNN. From the best of our knowledge, very little study includes non-contact facial of different conditions performing several physical exercises especially for fitness development application.

The remainder of this paper is structured as follows. Section 2 provides a synopsis of related work areas. Section 3 describes how to use CNN architecture to effectively estimate rPPG and measure HR and BR after subjects perform several physical exercises. Section 4 goes over the experimental setup and metrics for evaluating performance. Section 5 summarises the findings and discusses potential future work.

II. RELATED WORK

Previously, pulse oximetry was used to obtain non-invasive vital sign estimates by evaluating the PPG signal at different wavelengths. While non-invasive rPPG signal monitoring with a video camera is a viable technique for vital signs monitoring without even any electrodes or sensors directly contacting patients. The most of of rPPG estimation research was conducted under different illumination conditions, head movements, colour skin variations, and camera distance. All of these constraints were established using traditional methods. Deep learning has recently received a lot of attention.

A. Conventional Methods

The rPPG's early work is entirely based on a signal processing approach. To generate colour signals, skin detection is applied to a selected RoI and the average is tracked over time. The green channel of rPPG signals produced the strongest signal, which roughly corresponds to an oxyhemoglobin absorption peak. After that, the Blind Source Separation (BSS) technique is used to separate the rPPG signals from the noise [3].

The HR could be extracted using the Fast Fourier Transform (FFT), which also indicates the respiratory rate based on the RoI of the entire face. The independent component analysis (ICA) method predicts HR based on the largest spectral peak between 0.75 and 4Hz. Poh et al. [4] used viola-jones face detection to determine the mean intensity of the red (R), green (G), and blue (B) colour channels. Then, ICA was used to demix the pulse signal from the raw RGB signals. To handle skin-tone mismatch and subject motions, combined RGB channels must be normalized. Cheng et al. [5] use independent vector analysis to separate the facial components from the background regions in order to address illumination artefacts. Wang et al. [6] proposed modified projection orthogonal to the skin tone to extract pulse. Haan et al. [7] proposed CHROM method where the RGB channels were projected into the chrominance subspace to eliminate

motion components. Chen Lin et al. [8] proposed motion index method to develop a real-time contactless pulse rate status monitoring of head motion modelling using trajectories of tracked feature points. The BR in [9] was calculated using heart rate variability (HRV). Band-pass filtering and spectral analysis can be used to extract changes in the rPPG waveform. Similarly, in [10], the FFT is used to estimate HR from the rPPG signal, and motion analysis is used to estimate BR.

B. Deep Learning Methods

HR convolutional neural network (HR-CNN) detects regions of interest (RoI) in a pretrained Convolutional Neural Network (CNN) model, extracts de-noised signals, and predicts HR using an estimator. Spetlik et al. [11] proposed end-to-end HR estimation with a single scalar value of predicted HR as the output network. Qui et al. [12] created an EVM-CNN architecture that includes face detection, feature extraction, and estimation. The RoI defined the central part of the human face and formed spatial decomposition and temporal filtering based on 68 landmarks. The CNN was used to estimate HR based on these feature images. Chen and Mcduff [13] addressed subject motion issues using DeepPhys, a deep convolutional network. To estimate the HR signal, the network learns spatial masks and extracts features from blood volume pulse (BVP) data. Yu et al. [8] used deep spatiotemporal networks to reconstruct rPPG signals from raw facial videos as well. The measured rPPG signal has peaks that correspond to the R peak of ground truth electrocardiogram (ECG) signals. The other CNN model for HR estimation developed in [14] is trained using transfer learning on images constructed from synthetic rPPG signals. The synthetic rPPG signals are generated by interpolating BVP or ECG signals.

Another paper in [15] developed and trained a 2D CNN for skin segmentation on both skin and non-skin region samples. The detected skin region was then subjected to conventional rPPG algorithms (ICA and PCA) for HR estimation. Furthermore, the RoI colour information was extracted using the Generative Adversarial Network (GAN) architecture [16]. This method was used to create a high-quality noiseless rPPG signal in order to improve HR accuracy performance.

The research in [17] combines 2D CNN with Residual Neural Network (RNN) models such as Long Short-Term Memory (LSTM) to compare the performance with other HR-CNN models. The facial input was fed into a 2D CNN, which then extracted spatial features from RGB frames of video. In the context of temporal domain, RNN was used to form spatial features. Other than that, the Siamese-rPPG algorithm is based on Siamese 3D CNN [18]. The proposed framework is intended to overcome a variety of noises on various facial appearances. For a better pulse pixels extraction, the forehead and cheek regions were chosen as the RoI to extract significant rPPG information. The predicted rPPG signals were produced by fusion branches at the intermediate layer with two 1D convolutional layers, and average pooling layer. Table I shows the summary of related works using CNN for HR estimation based on facial videos. Most of the outcome was tested on public standard datasets like PURE, MAHNOB and COHFACE.

TABLE I. SUMMARY OF RELATED WORKS USING CNN

Ref.	Methods	MAE	RMSE
11	HR-CNN	10.09	13.14
12	EVM-CNN	6.85	6.95
13	DeepPhysPCA	2.35	4.50
16	Deep-HR	3.41	0.027

Instead of proposing a new framework to improve estimation performance through the use of deep learning-based methods, more understanding of CNN-based methods is required to clarify how it works with rPPG technology. Some research has also been conducted on the constraint and sensitivity of CNN-based networks in rPPG technology. According to the paper in [19], CNN for rPPG signal extraction is a learning-based information related to PPG signals, and the training is easily affected by the delay between the video data and the ground truth data. The CNN-based methods have some limitations, such as a limited number of frames. Some frameworks are not ideal for long-term signal estimation, possibly because they are only supporting for a specific dataset.

This paper addresses HR and BR estimation using the CNN method for subjects after they have completed a series of physical exercises in order to gain more valuable insights into the effectiveness of deep learning approaches in HR and BR prediction. Based on region-based skin detection, the recorded facial videos will be fed into CNN to create a spatio-temporal map of skin region. HR and BR are expected to be estimated using the spatiotemporal map. The comparison is made with ground truth data collected by contact-based HR and BR monitoring as presented in [20].

III. METHODOLOGY

This section will go into detail about several steps, such as preprocessing facial videos to extract skin regions and training spatio-temporal networks for rPPG estimation. Further investigation is required to calculate the HR and BR from our recorded dataset.

A. Preprocessing

The steps for preprocessing are as shown in Fig. 1. First, we convert the RGB input image to grayscale colour. Then, using Haar cascade for face detection, find the RoI in the input videos. The RoI of the face is divided into four sections: the Forehead, Eyes, Cheeks (including the Nose and Mouth), and Chin. The splitting steps are used to obtain local face features which capable to form sufficient RoI within the face area [22]. In Fig. 2, only two selected regions, the Forehead (RoI1) and both Cheeks and Nose (RoI2), serve as image training sequences for the CNN model. According to [23, 24], these areas typically contain more BVP information with the highest absorption region and are less affected by non-rigid motion such as smiling or eye blinking.

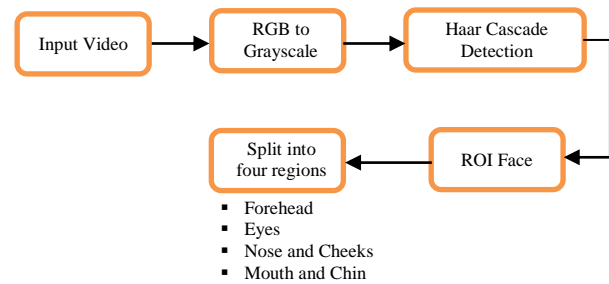


Fig. 1. Region based Skin Detection on Forehead, Cheeks, and Nose Areas.

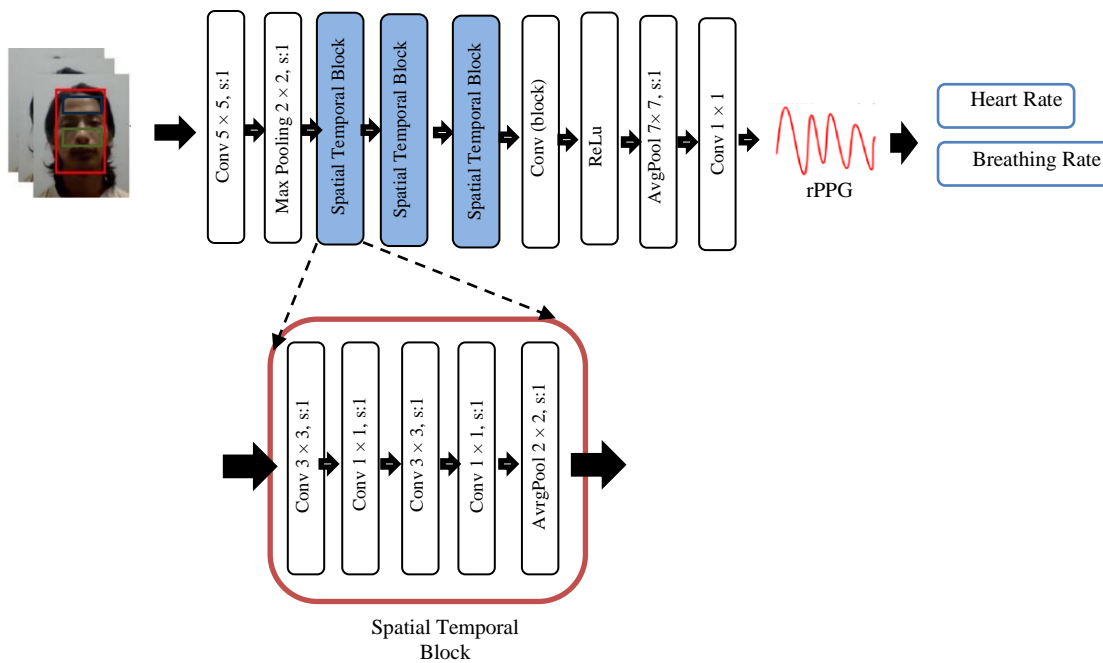


Fig. 2. The Overall Architecture from Region-based Skin Detection as an Input to the CNN and Spatio Temporal Convolutional Blocks. The Output is Single Scalar Vector of Predicted rPPG Signal.

B. Convolutional Neural Network

CNN was used specifically used to extract and estimate the single scalar vector of HR prediction. The CNN network is made up of multiple layers. Table II depicts the nine-layer CNN model structure. The first layer is a convolution kernel with a size of 5×5 and a stride of 1, which produces 64 feature maps. To reduce noise in extraction, the pool layer of the second layer used to perform maximum down-sampling of 2×2 . The next three layers are a spatio-temporal convolutional block with a 10×10 kernel in the convolution layer and a 2×2 kernel in the max-pooling layer. The high-level image features are generated by 200 feature maps. The next convolution layer has 500 feature maps, a kernel size of 4, and an average pooling layer with a 7×7 kernel. Convolution process, stride size, and pooling filter are used to minimise information loss while maintaining accurate physiological features for rPPG estimation.

The predicted rPPG signals are indicated by the output of spatio-temporal CNN from a single scalar vector. Each sample image was normalised to a size of 64×64 , with additional padding. Padding is required for all convolutions to maintain a consistent size. The non-activation ReLU function was used with a learning rate of 0.0005, a batch size of 200, and a training time period of 30.

C. Interbeat Interval

As shown in Fig. 3, peak detection is used to locate individual beats in the extracted rPPG signal. As a result, the inter-beat-interval (IBIs) is extracted from the rPPG signal. The IBIs are correspond to the time intervals between consecutive beats. Then, filters applied to the extracted IBIs to remove false positive/negative peak detections.

TABLE II. NETWORK ARCHITECTURE

CNN Layer	Layer Type	No. of Feature Map	Input Size
Layer 1	Conv.	20	64 x 64
Layer 2	mPool	20	32 x 32
Layer 3	ST Conv Block	200	16 x 16
Layer 4	ST Conv Block	200	8 x 8
Layer 5	ST Conv Block	200	8 x 8
Layer 6	ST Conv Block	200	4 x 4
Layer 7	ReLU	500	4 x 4
Layer 8	AvrgPool	16	1 x 1
Layer 9	Conv.	1	1 x 1

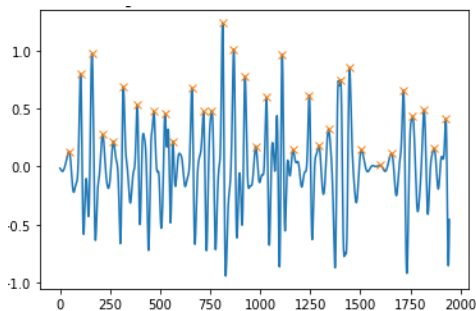


Fig. 3. Peak Detection of rPPG Signal.

D. Heart Rate Calculation

For valid peaks, the absolute IBI sequential difference should be less than 0.5 seconds. The HR is calculated by averaging all IBI over a time window and computing the inverse signals [21]. The IBI series is calculated as $IBI_{t_n} = t_n - t_{n-1}$ where t_n is the time of n -th detected peak. To simplify, $HR = 1/\overline{IBI}$ where \overline{IBI} is the mean of all inter-beat intervals within the time window. So, multiplying the HR in beats-per-minute by 60 yields the HR in beats-per-minute. For further analysis, the heart rate data is saved in a .csv file.

E. Respiration Rate Calculation

The normal respiration rate while people at resting condition is 12 to 20 beats per minute (bpm) [25]. Less than and more than this range are considered unstable respiration rates, indicating a slower (13bpm) or faster (>20bpm) breathing rate. Peak detection of the rPPG signal was used to estimate the BR using a time-domain technique. The duration of BR is defined as the time lag between the first and last detected peak. As a result, using spectral analysis on the estimated HR signal, the BR can be estimated. The power spectral density (PSD) was calculated, and the highest frequency amplitude within a plausible frequency range was chosen to represent the respiratory signal. The plausible respiratory frequency range was set from 0.1Hz to 0.4Hz.

F. Performance Metric Evaluation

The metrics evaluation used root mean square error (RMSE) and mean absolute error (MAE) to quantify the performance of proposed methods between predicted HR and BR rate and the ground truth. The fit standard error, or RMSE, was used to evaluate the best fitting of both estimated and ground truth data. As a result, as the RMSE decreases, so does the goodness of fit. The following equations were used to calculate the RMSE and MAE:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N |R_i - P_i|} \quad (1)$$

$$MAE = \frac{1}{N} |R_i - P_i| \quad (2)$$

Where R_i and P_i denote as the ground truth and predicted HR and BR, respectively. While N is the total number of heartbeats and respiration rate per minute.

IV. RESULT AND DISCUSSION

A. Datasets

This work used a self-collected dataset of 80 videos, as described in [20]. A total of 20 subjects with frontal view of face videos and 4 videos from each participant were obtained. All subjects are in good health and have signed a consent form to participate in the study. The video was shot under visible lighting with a webcam and a smartphone camera. Subjects perform three different conditions: 1) relaxed mode, (2) after a walking exercise, and (iii) after going up and down stairs. After each exercise, subjects were asked to sit still, and their facial expressions were recorded. The ground truth of each subject's HR and BR were taken within 60 seconds, with the pulse sensor and pulse oximeter collecting the data and

feeding it to the Arduino microcontroller for data evaluation. The final heartbeat values were displayed on the LCD screen.

Fig. 4 shows the schematic design of the prototype to collect the benchmark value of the HR and BR. The development and testing of this prototype has been done using standard PM100 pulse oximeter model. This prototype of self-collected facial dataset will be tested with another state-of-the-art algorithm to extract the rPPG signal using CNN-based model. The CNN network was implemented using Phyton with Tensor Flow and Keras framework within Colab notebook. The proposed work is validated with our self-collection post-exercise dataset which contains facial videos of different lighting conditions and taken from smartphone or webcam. We have produced three different set of exercise mode to evaluate the proposed framework efficiency. At the same time, validate the effectiveness of our health monitoring prototype. We observed that the dataset which consists varying pose of less motion yields better accuracy in terms of HR and BR estimation. Motion artefacts and lighting variations also causing lots of noise in the extracted rPPG signal. Hence, the number of false positive or negative peak detections will increase.

B. Heart Rate Analysis

The simulation work was done in order to compare the measured HR to the predicted HR. The comparisons for the rest and post-exercise conditions are shown in the following figures. The efficacy of the proposed RoI was experimentally validated using the extracted rPPG signal for both HR and BR estimation. Fig. 5 depicts a comparison of measured and estimated HR when the subject is at rest. The estimated HR falls within the normal range of 60 to 100bpm. Fig. 6 and 7 depict a sample of the subject after exercises, which are walking and staircase exercises. The difference between measured and estimated HR for post-exercise falls within a range of more than 80 bpm. According to our observations, the estimated HR based on the CNN model agrees well with the measured HR from the prototype developed in our previous work. Our proposed method accurately calculates the changes in estimated HR from rest to post exercise.

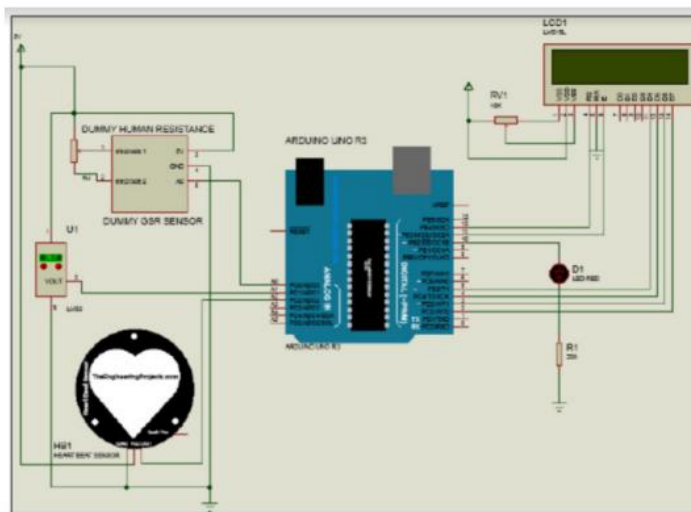


Fig. 4. The Schematic Design of the Health Monitoring Prototype [20].

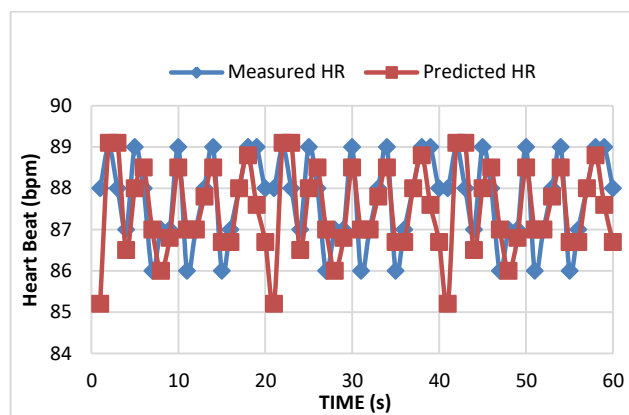


Fig. 5. Measured Heart Rate against the Estimated Heart Rate while Subjects Resting Conditions.

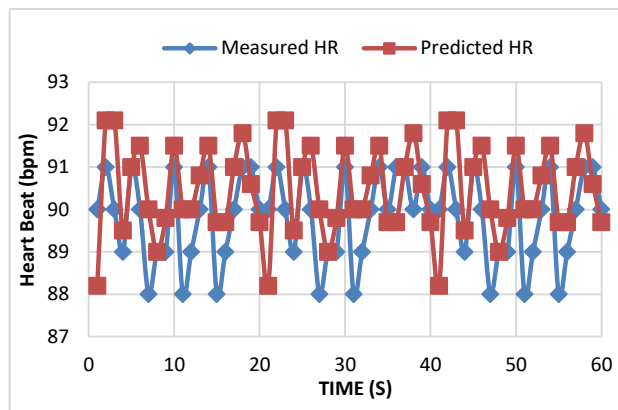


Fig. 6. Measured Heart Rate against the Estimated Heart Rate for Subject after Walking Exercise.

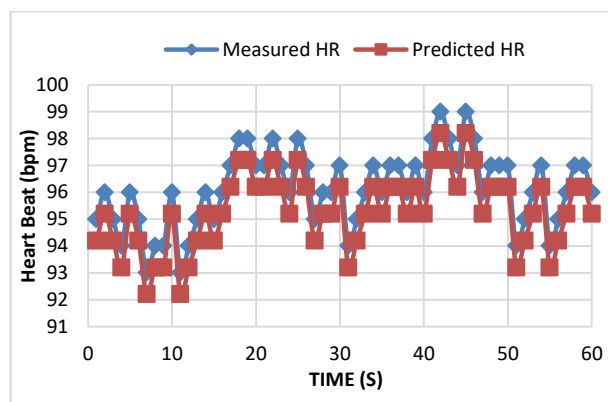


Fig. 7. Measured Heart Rate against the Estimated Heart Rate for Subject after Stairs Exercise.

From the results in Table III, the RMSE and MAE are derived from the average measured of the actual HR and predicted HR for all subjects for several types of exercises. The average of actual HR for resting, walking and after stairs exercise are 87.75, 89.8 and 95.3, respectively. The MAE for our proposed method using our self-collected physical exercise dataset is very good, falling within the 2.16 ± 2.2 beats/min range. To evaluate our dataset using our proposed method, we compared it to other state-of-the-art methods based on CHROM, ICA, PCA and HR-CNN. Our proposed

method outperforms the traditional CHROM, ICA and PCA methods with low MAE and RMSE. We conclude that CNN based methods are more accurate and present relevant beats from the pulse rate series for HR estimation. We demonstrate the similar process for BR performance.

TABLE III. HR ANALYSIS BASED ON AVERAGED MEASURE, RMSE AND MAE

Method	Exercise	Predicted _{avg} HR	RMSE	MAE
CHROM[7]	Rest	75.97	11.47	5.4
	Walk	78.98	10.92	7.64
	Stairs	69.6	21.19	9.48
ICA[15]	Rest	75.2	5.22	0.97
	Walk	67.65	12.67	8.21
	Stairs	68.8	14.94	10.05
PCA[15]	Rest	85.43	9.72	3.9
	Walk	77.88	2.17	1.14
	Stairs	79.03	7.44	7.98
HR-CNN[11]	Rest	83.05	2.52	2.80
	Walk	80.65	2.55	2.91
	Stairs	88.37	2.37	2.15
Proposed Method	Rest	87.47	1.92	1.72
	Walk	80.48	2.37	2.96
	Stairs	86.1	2.64	1.8
	Overall		2.31	2.16

C. Breathing Rate Analysis

Fig. 8 depicts a comparison of measured and estimated BR when the subject is at rest. The estimated BR falls within the normal range of 15 to 20bpm. Fig. 9 and 10 show a sample of the subject after exercises, which are walking and stairwell exercises. The difference between measured and estimated BR for post-exercise falls within a range of more than 20 bpm. The curves are simulated from the extracted rPPG signals within a constant distance from the camera. ow SNR. Rapid changes of breathing rate can be detected from resting position and stairs exercise. This is accomplished by selecting the peak in the spectrum that provides the highest SNR for the pulse signal.

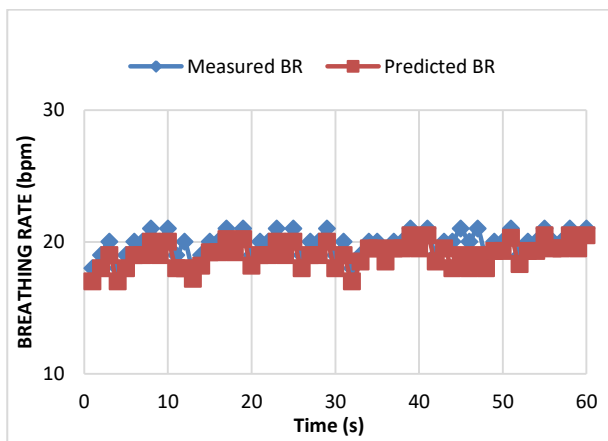


Fig. 8. Measured Heart Rate against the Estimated Heart Rate while Subjects at Resting Conditions.

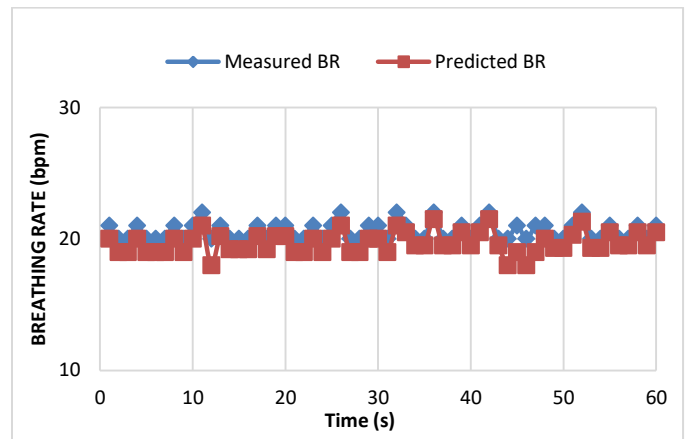


Fig. 9. Measured Heart Rate against the Estimated Heart Rate for Subject after Walking Exercise.

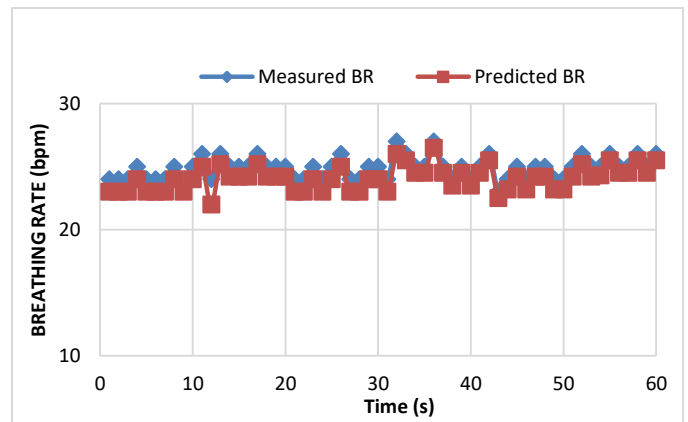


Fig. 10. Measured Heart Rate against the Estimated Heart Rate for Subject after Stairs Exercise.

Table IV derived the RMSE and MAE from the average measured of the actual BR and predicted BR for all subjects for several types of exercises. The average of actual BR for resting, walking and after stairs exercise are 19.9, 20.5 and 24.8, respectively. The MAE for our proposed method of estimating BR is between 2.16 and 2.2 beats/min. We demonstrated the effectiveness of BR peak finding interpolation and reported an acceptable MAE of less than 2 bpm. The method is compared to various state-of-the-art methods (CHROM, PCA, and ICA) and an HR-CNN-based method. For our specific post-exercise dataset, our proposed technique produced valuable findings. Although the use of the CNN framework in our suggested method does not compete with other deep learning methods, it is sufficient to examine the practicality of our prototype as a non-invasive health assessment system.

The results presented for HR and BR estimation show a high level of agreement between calculated and ground truth measurements. Based on the HR and BR analysis in Table III and Table IV, the CHROM method has the highest RMSE and MAE. CHROM is typically used to remove noise caused by light reflection via colour difference channel normalisation. For the purposes of comparison, both the ICA and PCA methods are types of BSS techniques in which independent signal sources are filtered from a mixed signal. Both ICA and

PCA improved the RMSE and MAE in this simulation work. Most rPPG research studies have a very limited post-exercise dataset. Despite the fact that several deep learning methods have been implemented in the facial-based rPPG dataset, our results validated on this self-collected post-exercise dataset are comparable to other state-of-the-art algorithm.

Based on the experimental results and performance evaluation, the proposed RoI of facial regions is effective for CNN model in forming spatio-temporal features for rPPGg signal estimation. A low percentage of MAE and RMSE indicates that the RoI selection is significant in improving the SNR for better signal quality assessment. We can assume that the majority of significant pulse pixels are located on forehead and cheeks. In practice, the most significant impediment to accurate HR and BR analysis is false peak detection. The signal was interpolated at 256 Hz to sharpen the peaks and manage the latency. Beat-to-beat pulse rate values were computed from the interbeat intervals. When calculating the interval between beats, a false peak detection appears as an incorrect beat and causes a major error in the HR and BR analysis of a healthy person. Removing unnecessary peaks may also have an impact on HR and BR estimation.

However, there are few limitations included in our study. The HR and BR estimation from the extracted rPPG signal requires further evaluation using CNN. The end-to-end approach of CNN model will be efficiently incorporated into our model's future development. Due to the limitations of the dataset used for this study, we will increase the duration of exercise and the recorded time to produce more variation in the HR and BR signal patterns. In the future, we will use adaptive RoI detection to improve the efficacy of our region-based skin detection model.

TABLE IV. BR ANALYSIS BASED ON AVERAGED MEASURE, RMSE AND MAE

Method	Exercise	Predicted _{avg} BR	RMSE	MAE
CHROM[7]	Rest	17.28	14.98	4.46
	Walk	17.85	7.99	3.46
	Stairs	22.30	13.77	2.38
ICA[15]	Rest	18.28	2.16	1.01
	Walk	18.85	3.17	1.01
	Stairs	23.3	2.95	3.93
PCA[15]	Rest	18.73	4.08	2.97
	Walk	19.30	5.09	2.97
	Stairs	23.75	2.87	1.89
HR-CNN[11]	Rest	18.52	2.53	1.89
	Walk	20.25	1.94	1.85
	Stairs	25.61	1.52	1.51
Proposed Method	Rest	19.03	2.93	0.89
	Walk	19.60	0.94	1.89
	Stairs	24.05	1.72	1.81
	Overall		1.86	1.53

VI. CONCLUSION AND FUTURE WORK

We may conclude that the results are promising enough to support the effectiveness of the CNN model in extracting relevant pulse pixels for further analysis. For HR estimation, our methods achieved an average RMSE of 2.31 and MAE of 2.16. The overall BR estimation had an average RMSE of 1.86 and an MAE of 1.53. Our methods perform well in a variety of post-exercise facial video streams under controlled lighting conditions, whether utilizing a camera or a smartphone.

In the future, an end-to-end CNN approach will be established using this dataset for simulating HR and BR estimation. This system can be upgraded to detect sudden changes in heart rate and breathing patterns. Furthermore, this system made use of a partial face region, which was expected to contribute the most blood variation. The system will be improved in real-time scenario especially when subject performing head movements.

Other factors under consideration for future research include increase the number of hidden layers in the CNN framework and optimizing the network design to achieve more promising outcomes. The number of layers, number of filters per convolutional layer, and number of neurons per dense layer could all have a significant impact and provide an automatic method of determining the best network architecture. Further investigation can be performed such as improving the CNN framework and comparing the influence of colour channel performance, particularly in terms of rPPG signal accuracy and artifact removal.

ACKNOWLEDGMENT

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 Grant (Vot H756).

REFERENCES

- [1] Jean Pierre Tincopal , Paulo Vela-Anton, Cender U. Quispe-Juli, Anthony Arostegui, "Development of an IoT Device for Measurement of Respiratory Rate in COVID-19 Patients", In International Journal of Advanced Computer Science and Applications, pp. 1009-1016, Vol. 13, No. 4, 2022.
- [2] Chun-Hong Cheng, Kwan-Long Wong, Jing-Wei Chin et al., "Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda", In Sensors (Basel, Switzerland), Vol. 21, No. 18, 2020.
- [3] Ming-Zher Poh, Daniel J. McDuff, Rosalind W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation", In Optics Express, pp. 10762-10774, Vol. 18, No. 10, 2010.
- [4] Poh, M.Z., McDuff, D.J., and Picard, R.W, "Advancements in noncontact, multiparameter physiological measurements using a webcam", In IEEE Transactions on Biomedical Engineering, pp. 7-11, Vol. 58, No. 1, 2010.
- [5] Juan Cheng, Xun Chen, Lingxi Xu. et al., "Illumination Variation-Resistant Video-Based Heart Rate Measurement Using Joint Blind Source Separation and Ensemble Empirical Mode Decomposition," In IEEE Journal of Biomedical and Health Informatics, pp. 1422-1433, Vol. 21, No. 5, 2017.
- [6] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk et al., "Algorithmic Principles of Remote PPG," In IEEE Transactions on Biomedical Engineering, pp. 1479-1491, Vol. 64, No. 7, 2017.
- [7] Gerard de Haan, Vincent Jeanne, "Robust Pulse Rate from Chrominance-Based rPPG," In IEEE Transactions on Biomedical Engineering, pp. 2878-2886, Vol. 60, No. 10, 2013.

- [8] Yu-Chen Lin, Nai-Kuan Chou, Guan-You Lin et al., "A Real-Time Contactless Pulse Rate and Motion Status Monitoring System Based on Complexion Tracking," In *Sensors (Switzerland)*, Vol. 17, No. 7, 2017.
- [9] Ming-Zher Poh, Daniel J. McDuff, Rosalind W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," In *IEEE Transactions on Biomedical Engineering*, pp. 7-11, Vol. 58, No. 1, 2011.
- [10] Giovanni Cennini, Jeremie Arguel, Kaan Aksit et al., "Heart Rate Monitoring Via Remote Photoplethysmography with Motion Artifacts Reduction," In *Optics Express*, pp. 4867-4875, Vol. 18, No. 5, 2010.
- [11] Radim Spetlik, Jan Cech, Vojtěch Franc et al., "Visual Heart Rate Estimation with Convolutional Neural Network," In the *British Machine Vision Conference*, 2018.
- [12] Ying Qiu, Yang Liu, Juan Arteaga-Falconi et al., "EVM-CNN: Real-Time Contactless Heart Rate Estimation from Facial Video," In *IEEE Transactions on Multimedia*, pp. 1778-1787, Vol. 21, No. 7, 2019.
- [13] Weixuan Chen, Daniel McDuff, "Deepphys: Video-Based Physiological Measurement Using Convolutional Attention Networks," In *Proceedings of the European Conference on Computer Vision*, pp. 349-365, 2018.
- [14] Rencheng Song, Senle Zhang, Chang Li et al., "Heart Rate Estimation from Facial Videos Using a Spatiotemporal Representation with Convolutional Neural Networks," In *IEEE Transactions on Instrumentation and Measurement*, pp. 7411-7421, Vol. 69, No. 10, 2020.
- [15] Chuanxiang Tang, Jiwu Lu, Jie Liu, "Non-contact Heart Rate Monitoring by Combining Convolutional Neural Network Skin Detection and Remote Photoplethysmography via a Low-Cost Camera," In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1309-1315, 2018.
- [16] Mohammad Sabokrou, Masoud Pourreza, Xiaobai Li, Mahmood Fathy, Guoying Zhao, "Deep-HR: Fast Heart Rate Estimation from Face Video Under Realistic Conditions," In *Expert System with Applications*, 2020.
- [17] Xingjian Shi, Zhoung Chen, Hao Wang et al, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," In *Proc. of 28th International Conference on Neural Information Processing Systems*, pp. 802-810, Vol. 1, 2015.
- [18] Yun-Yun Tsou, Yi-An Lee, Chiou-Ting Hsu et al, "Siamese-rPPG network: remote photoplethysmography signal estimation from face videos," In *proceedings of the 35th Annual ACM Symposium on Applied Computing*, pp. 2066-2073, 2020.
- [19] Zhan Q, Wang W, de Haan G, "Analysis of CNN-based remote-PPG to understand limitations and sensitivities," In *Biomedical Optics Express*, pp. 1268-1283, Vol. 11, No. 3, 2020.
- [20] Nor Surayahani Suriani, Nur Adlina Jumain, Abdalla Abdurahman Ali, "Facial Video based Heart Rate Estimation for Physical Exercise," In *IEEE Symposium on Industrial Electronics & Applications*, pp.1-5, 2021.
- [21] Azizullah Kakar, Naveed Sheikh, Bilal Ahmed, Saleem Iqbal, Abdul Rahman, Saboor Ahmad Kakar, Arbab Raza, Samina Naz, Junaid Baba, "Systematic Analysis and Classification of Cardiac Rate Variability using Artificial Neural Network," In *International Journal of Advanced Computer Science and Applications*, pp 746-750, Vol.9, No.11, 2018.
- [22] Dae-Yeol Kim, Kwangkee Lee, Chae-Bong Sohn, "Assessment of ROI Selection for Facial Video-Based rPPG," In *Sensors (Basel)*, ISSN: 1424-8220, pp. 7923, Vol. 21, No. 23, 2021.
- [23] D. Datcu, M. Cidota, S. Lukosch and L. Rothkrantz, "Noncontact automatic heart rate analysis in visible spectrum by specific face regions", In *ACM International Conference Proceeding Series*, Vol. 767, 2013.
- [24] S. Kwon, J. Kim, D. Lee and K. Park, "ROI analysis for remote photoplethysmography on facial video", In *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4938-4941, 2015.
- [25] J. Chen and A. Flabouris, "Respiratory Rate: The Neglected Vital Sign," In *The Medical Journal of Australia*, Vol.188, No. 11, pp. 657, 2008.