

Contactless Algorithm for Identifying the Facial Area Containing the Most Vital Sign-Related Information Using Machine Learning.

Moussa Mmadi¹, George N. Kamucha², Ciira wa Maina³

¹Department of Electrical Engineering, Pan African University Institute for Basic Sciences, Technology and Innovation, Nairobi, Kenya

²Department of Electrical and Information Engineering, University of Nairobi, Nairobi, Kenya

³Department of Electrical Engineering, Dedan Kimathi University of Technology, Nyeri, Kenya

¹Corresponding Author: moussa.mmadi@students.jkuat.ac.ke

Abstract

Introduction: In recent years, there have been significant advancements in the field of biological parameter measurement. The emergence of contactless technology now enables the extraction of vital sign data using a standard personal computer (PC) equipped with a webcam. This system captures facial footage to obtain a remote photoplethysmography (rPPG) signal.

Objectives: Identifying the correct extraction area, known as the Region of Interest (ROI), is a crucial part of this process. Traditionally, the forehead is used as the primary site for extracting physiological parameters. However, this method poses challenges when patients have injuries or fractures on the forehead that need to be covered. To overcome this issue, we developed an automated computer vision model utilizing a machine learning technique that can analyse and classify blood flow quality in real-time across three facial ROIs: the forehead, right cheek, and left cheek.

Methods: The procedure starts by identifying the subject's face in the first frame and ensuring it remains tracked throughout the video. Subsequently, three mask regions are created to facilitate feature extraction using the bitwise operation method. In the final step, a Support Vector Machine (SVM) model is used to ascertain which of the three regions holds the most crucial physiological information.

Results: The proposed system shows promising results when evaluated using the PURE testing dataset. It delivered outstanding performance, achieving an accuracy rate of 99% and a False Positive Rate (FPR) of merely 0.7%.

Conclusions: The suggested approach is highly efficient in monitoring the forehead, right cheek, and left cheek of a patient's face, identifying the region with the richest physiological data among them. This method was created and evaluated with 10 participants who were seated in a chair, facing the camera. The camera was positioned 1 meter away from the subjects.

Keywords: non-contact; computer vision; machine learning; photoplethysmography.

1. Introduction

Monitoring the physiological condition of the human body often involves tracking key vital signs, including heart rate (HR), body temperature (BT), respiratory rate (RR), blood oxygen saturation (SpO₂), heart rate variability (HRV), and blood pressure (BP) [3–7]. These indicators are valuable tools for assessing an individual's physical well-being, identifying potential illnesses, and tracking recovery progress. Long-term monitoring of

physiological parameters provides useful insights into a patient's condition, and in some cases, is essential for diagnosing certain conditions, such as ectopic heartbeats [4].

Photoplethysmography (PPG) is an economical and non-invasive technique for assessing physiological parameters [2–6]. The fundamental concept of PPG involves employing a light source alongside a photodetector to assess variations in the blood vessel volume beneath the skin. As the light source

illuminates the tissue, the photodetector captures minor fluctuations in the light's reflection or transmission intensity due to blood flow, resulting in the generation of the PPG signal [7]. Devices like pulse oximeters and fitness trackers utilize PPG to non-invasively monitor minor skin changes based on this principle. Nonetheless, these conventional contact devices have several drawbacks, such as being inappropriate for identifying skin conditions in sensitive groups like infants and individuals with skin disorders [8]. They can also cause discomfort or even lead to skin infections with prolonged use [9], and their accuracy can be compromised by factors like skin moisture, temperature, color, and patient movement [10].

To circumvent these drawbacks, researchers have started investigating non-contact techniques for remote heart rate monitoring, with rPPG emerging as a formidable alternative. This method employs a camera, such as a webcam, infrared camera, or RGB camera, to capture video footage of a person's face, detecting minute color variations in the skin to produce the remote PPG signal [11]. The underlying principle of rPPG is akin to that of traditional PPG, where the pulsation of blood through the cardiovascular system alters the blood volume in the microvascular tissue beneath the skin with each heartbeat, creating periodic waves. Presently, rPPG is recognized as superior because it eliminates the need for subjects to wear contact devices, thus avoiding their associated disadvantages, and it is also ideal for prolonged continuous monitoring, accommodating various patients. Additionally, the camera necessary for the rPPG technique is inexpensive and readily accessible, making it highly suitable for extensive promotion and use [1]. However, employing rPPG techniques in practical settings is more challenging due to factors such as lighting variations, facial hair, and skin tone, all of which can affect the precision of the signal obtained. Additionally, the rPPG signal is inherently weaker compared to that derived from traditional contact methods because of fundamental differences, necessitating meticulous and accurate processing.

To overcome these challenges, obtaining a precise rPPG signal is essential. Consequently, a preprocessing algorithm is required to identify a

region with distinct blood perfusion. Given the limited research on analysing blood perfusion through image processing, our study concentrates on these elements. The primary aim of this paper is to present a computer vision model that utilizes a machine learning technique for real-time analysis, enabling the classification of a patient's blood flow quality across the ROIs: the forehead, right cheek, and left cheek.

The key contributions are as follows: i) A comprehensive literature review on the extraction and analysis of the rPPG signal is performed. ii) An image processing algorithm is developed to extract features related to blood perfusion, and its performance is evaluated under different lighting conditions and recording setups. iii) A machine learning model is designed to classify the blood flow quality (Good or Poor) for each ROI.

The remainder of the work is structured as follows. Section 2 gives an overview of the most widespread image processing approaches for rPPG extraction from videos. Section 3 provides the methodology that has been used for tracking the ROIs and extracting features linked to the blood perfusion. Sections 4 and 5 give the results and the conclusions, respectively.

2. Related work

As previously mentioned, rPPG involves several techniques. Below, we offer a summary of the key methods in image processing for rPPG.

Although rPPG can be applied to all skin areas, most research focuses on the face due to its several benefits over other skin regions. Primarily, the face is usually not obscured by clothing. Additionally, it has a large skin surface and excellent blood flow, which are crucial for pulse signal extraction. A fundamental method involves using a cascade classifier to identify the face, as outlined by Viola and Jones [13]. This algorithm employs Haar-like image filters to detect facial features, resulting in a bounding box that indicates the face's position and size. Pulse signals for each color channel are then derived by averaging the image data within the face's bounding box [14]. A significant limitation of this technique is its restriction to specific viewing angles of the face. To address issues of rotation and subject movement, a widely adopted solution is the

implementation of a point tracking algorithm, such as the Kanade-Lucas-Tomasi Tracker [15], which adjusts the bounding box to follow the subject's movements [16], [17].

In [18], a method for tracking ROI was introduced, which involves generating regional proposals and utilizing multi-scale ROIs. Reference [19] presented a dynamic ROI tracking technique capable of monitoring minor movements and variations in lighting, with the ability to operate in real time. Wei et al. [20] employed the MediaPipe Face Mesh3.1 framework [21] to identify 468 facial landmarks within a frame and dynamically track the face region. Similarly, Pagano et al. [22] utilized the MediaPipe Face Mesh, integrating it with a support vector machine to extract the region of interest. Once a face is detected, it is often necessary to exclude areas with lower signal-to-noise ratio (SNR), such as the eyes and mouth, which can exhibit significant fluctuations during typical human activities like blinking and speaking. Wong et al. [23] introduced a novel 'angle map' representation of facial regions to identify areas with a higher SNR.

In earlier research, Verkrusse et al. were the pioneers in suggesting the use of consumer-grade cameras to derive rPPG signals for heart rate monitoring [10]. Their study revealed that the Red, Green, and Blue (RGB) signal's different channels exhibited varying levels of PPG signal strength, with the green channel showing the most pronounced pulsatile signal.

3.2.

Poh et al. [14] introduced an Independent Component Analysis (ICA) algorithm that utilizes joint approximate diagonalization of eigenmatrices to eliminate correlations and higher-order dependencies among RGB channels, thereby extracting HR components in both stationary and naturally moving sitting scenarios. The root mean square error (RMSE) for the motion scenario was reduced from 19.36 bpm to 4.63 bpm, indicating the potential of ICA for HR estimation.

3. Methods

Figure 1 illustrates the system diagram. The method we propose takes the facial RGB color from video footage of subjects as input and produces an assessment of blood flow quality across three designated regions of interest. Initially, the process

involves detecting the subject's face in the initial frame and maintaining its tracking throughout the video. The next step focuses on generating three mask regions, which will be utilized to extract features through the bitwise operation technique. Finally, an SVM model is employed to determine which of the three regions contains the most significant physiological data. Each of these steps will be elaborated upon in the following subsections.

Recording the video and face detection

Face detection is a crucial component in numerous applications, including face authentication, recognition, tracking, and emotion recognition. The purpose of a face detection algorithm is to ascertain whether a face is present in an image. While humans can easily identify faces, this task poses a significant challenge for computers. Consequently, it has been a fascinating area of research for many years. In video processing, detecting a subject's face from a video frame is vital. To achieve accurate face detection, a dependable face tracking method is necessary. Therefore, for this project, the 'CascadeObject-Detection' method [24], [25] from the Open Source Computer Vision (OpenCV) library, developed by Viola-Jones [26], was employed. In this part, the subject will face the webcam, and a video recording will be carried out. During the recording, the camera will try to detect the face and save an image of it.

Detection of the ROIs

To determine this ROI, three rectangular masks were applied to the facial image identified in the earlier section. This method can identify the following three facial areas: the Forehead and both Cheeks. Constructing a mask within an image typically involves creating a binary mask where pixels within the desired region are assigned a value of one (or True=1), while those outside are given a value of zero (or False=0). These masks match the dimensions of the image captured by the camera, and mathematically, they are represented as [31]:

$$R_m(x, y) = \begin{cases} 1, & \text{if } x_1 \leq x \leq x_2 \text{ and } y_1 \leq y \leq y_2 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

With (x_1, y_1) the coordinate at the top-left corner, and (x_2, y_2) the coordinate at the bottom-right

corner, and (x, y) is the set of points belonging to the region of interest of R_m

Left_Cheek_mask: When applied to the image, it results in an output where every area is blacked out except for the left cheek.

Right_Cheek_mask: Upon application to the image, it produces an output where all areas are black except for the right cheek.

Forehead_mask: Similarly, this mask can be used to obscure all facial regions except the forehead.

After preparing all the masks, the subsequent step involves applying these masks to the facial image through a method known as bitwise operation, a common technique in image processing. Bitwise operations manipulate the binary form of pixel values in images, allowing pixels to be processed as binary rather than numerical values. This approach

Various bitwise operations can be employed in image processing, including:

- **AND:** The bitwise AND operation compares corresponding bits between two images. It produces an image where the pixel value is 1 if both input images have a pixel value of 1, and 0 otherwise.
- **OR:** The bitwise OR operation evaluates the corresponding pixels in both input images. The resulting image has a pixel value of 1 if either of the input images has a pixel value of 1, and 0 otherwise.
- **XOR:** The bitwise XOR operation compares the corresponding bits of two binary input images. It outputs an image with a pixel value of 1 when the bits differ in the input images, and 0 when they are the same.
- **NOT:** The bitwise NOT operation inverts the bits of an image, changing each bit from 1 to 0 and vice versa.

The purpose of using bitwise is to apply a mask on the original image to hide the face and only reveal the part that we need for future processing. According to our needs, the bitwise AND operation has been chosen for use because it aligns with our purpose. The result can be seen in section 4 of this paper.

3.3. Extracting the features

is particularly beneficial for tasks like image masking, where selecting a specific part of an image based on another image's values is required [30].

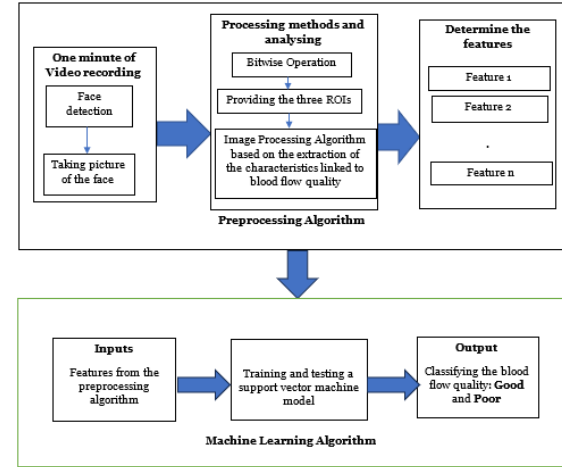


Figure 1. General methodology for blood flow classification.

To select the optimal region from these three ROIs, it is crucial to assess a range of parameters related to both image features and physiological attributes. From [29]-[33], we learnt that among the parameters that should be taken into account are:

3.3.1. Variance and standard deviation

Detecting changes in pixel intensity can reveal details about texture and the dynamics of blood flow. The formulas for variance and standard deviation are provided as [33]:

- Variance σ^2 of the pixel intensities in the region of interest are given by:

$$\sigma^2 = \frac{1}{N} \sum_{(x,y) \in ROI} (I_{x,y} - \mu)^2 \quad (2)$$

With $I_{x,y}$ is the pixel value at the coordinate (x, y) , and N The number of pixels in the ROI. The expression $(I_{x,y} - \mu)^2$ presents the squared difference between each pixel value and the mean pixel value.

μ represents the mean pixel value in the ROI, and it is defined in [33] as:

$$\mu = \frac{1}{N} \sum_{(x,y) \in ROI} I_{x,y} \quad (3)$$

- The standard deviation σ_n of the noise in the ROI is given by:

$$\sigma_n = \sqrt{\frac{1}{N} \sum_{(x,y) \in ROI} (I_{x,y} - \mu)^2} \quad (4)$$

3.3.2. Entropy

Entropy quantifies the complexity of texture, which might suggest irregularities in perfusion. As noted in [33], entropy is defined as:

$$Entropy = -\sum_i P(I_i) \log_2 P(I_i) \quad (5)$$

Where $P(I_i)$ It is the probability of pixel intensity. And mathematically it is given as:

$$P(I_i) = \frac{h(I_i)}{N} \quad (6)$$

With $h(I_i)$ is the number of pixels with intensity I_i

And N The total number of pixels in the ROI.

3.3.3. Gaussian filter

To smooth the image, eliminate noise, and emphasize significant intensity changes associated with blood flow, a Gaussian filter is utilized. The Gaussian equation is represented as [31]:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (7)$$

(x, y) The coordinates of a pixel relative to the center of the filter. And σ is the standard deviation of the Gaussian distribution.

3.3.4. Gabor filter

Gabor captures information about frequency and orientation, which aids in identifying structural patterns in perfusion. The Gabor filter is employed for analyzing edges and textures. It is characterized as [32]:

$$G(x, y; \lambda, \theta, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right) \quad (8)$$

Where:

$$x' = x \cos \theta + y \sin \theta \quad (9)$$

$$y' = -x \sin \theta + y \cos \theta \quad (10)$$

λ is the wavelength, θ is the orientation, σ is the standard deviation of the Gaussian envelope, γ is the spatial aspect ratio, and ϕ is the phase offset.

3.3.5. signal-to-noise ratio

In the realm of image processing, determining the signal-to-noise ratio of an image is a crucial step. This evaluation helps us gauge the image's quality. The SNR is particularly useful in scenarios like

evaluating image quality and diagnostic precision. In such instances, a higher signal-to-noise ratio indicates superior image quality, which is essential for accurate interpretation. It quantifies the usefulness of the signal relative to the noise. The signal-to-noise ratio is defined as the ratio of the average signal value to the noise's standard deviation.

$$SNR = \frac{\mu}{\sigma_n} \quad (11)$$

With μ and σ_n have been calculated respectively in equations (3) and (4).

3.4. Using machine learning for detecting the best region of interest

The main purpose of this part is to classify the blood flow quality (good or poor) in the different regions (forehead, right cheek, and left cheek). However, to achieve this objective, a machine learning algorithm based on classification was used.

3.4.1. Dataset

This study utilizes the PURE dataset [29]. To access it, it is possible through [34], which includes data from 10 individuals engaged in various controlled head movements in front of a camera. Throughout these sessions, both the image sequences of the head and reference pulse measurements were captured. The group of 10 participants (comprising 8 males and 2 females) was recorded across 6 distinct setups, resulting in a total of 60 sequences, each lasting 1 minute.

3.4.2. Labelling the dataset

To categorize blood flow quality into Good and Poor, it is essential to establish labels based on the features extracted. A detailed explanation using a structured method is presented in Table 1. To determine numerical thresholds for each quality label (Good and Poor), a statistical analysis-based method in [29]-[31] was applied to the dataset. The following is an approach for setting these thresholds: To compute feature distribution, we calculate the Mean (μ) and Standard Deviation (σ) for each feature throughout the dataset. A typical method involves establishing quality labels using percentiles. The findings are presented in Section 4, Figure 5.

3.4.3. Evaluation Metrics

The proposed system model was assessed using standard metrics from the literature, including accuracy (ACC), precision (PR), recall (R), F1-score, and false positive rate (FPR). Accuracy measures the model's rate of correct classifications, precision indicates the positive predictive value, recall (also known as sensitivity) reflects the True positive rate (TPR), and the F1-score is calculated as the harmonic mean of precision and recall. The equations for each of these metrics are defined in [22] as follows:

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (12)$$

$$PR = \frac{TP}{(TP + FP)} \quad (13)$$

$$recall = \frac{TP}{(TP + FN)} \quad (14)$$

$$FPR = \frac{FP}{(FP + TN)} \quad (15)$$

$$F1 - score = 2 * \frac{(PR * Recall)}{(PR + Recall)} \quad (16)$$

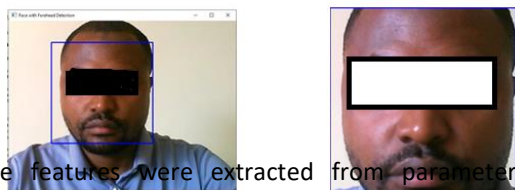
TP represents True Positives, which are the correctly predicted positive cases. TN stands for True Negatives, indicating the correctly predicted negative cases. FP refers to False Positives, where positive cases are incorrectly predicted. FN denotes False Negatives, where negative cases are incorrectly predicted.

4. Results

This section provides a detailed, step-by-step presentation of the experimental results, as outlined in the methodology.

4.1. Video and face detection

In this part, the subject faces the webcam, and a video is recorded. During the recording, the camera detects in real-time the face and saves an image of it. The result is shown in Figure 2.



The features were extracted from parameters including SNR, average intensity, variance, standard deviation, entropy, Sobel filter, Gaussian filter, and

(a) (b)

Figure 2. (a) is recording video with face detection. (b) is the image's face detected

The purpose of that picture (b) is to carry out a pre-processing on it, allowing us to decide on which area of the face will be extracted for the rPPG signal. All steps concerning this pre-processing are explained in the next subsection.

4.2. Extracting the ROIs

Figure 3 presents the three masks that have been developed. Figure 4 shows the result of the process of extracting the ROIs. The technique consisted of using the bitwise AND operation between the original image in Figure 2 (b) and the three masks in Figure 3.

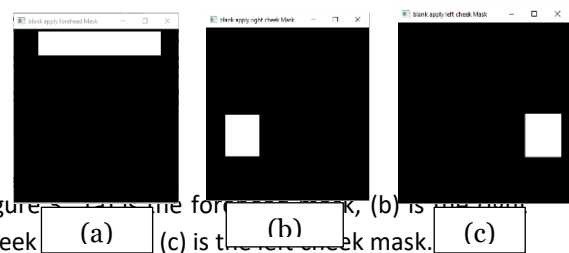


Figure 3. (a) is the forehead mask, (b) is the right cheek mask, (c) is the left cheek mask.

Machine learning algorithm

In this section, a support vector machine was employed to categorize the quality of blood flow (either good or poor) across various regions, including the forehead, right cheek, and left cheek. Initially, a preprocessing step was implemented, which relied on image processing techniques. Specifically, the algorithm accessed images from the PURE database.

Through our preprocessing method, each image underwent face detection, and the three regions of interest were extracted. The subsequent step involved constructing the dataset. This dataset incorporated features that pertain to both image and physiological characteristics within each ROI.

Gabor filter. The final phase involved labeling the dataset by assessing the quality of blood flow,

categorizing it as Good if the ROI demonstrated high quality, and Poor if the quality was low.

To assign these labels, numerical thresholds for each quality category (Good and Poor) were established using a statistical analysis method. The approach of defining Good Blood Flow (High Quality) and Poor Blood Flow (Low Quality) through statistical thresholds, such as μ (mean) $\pm 0.5\sigma$ (standard deviation), is a heuristic method commonly used in image processing and signal analysis [31]. When a feature (e.g., variance, standard deviation, entropy) significantly surpasses

the mean ($\mu + 0.5\sigma$), it indicates high texture complexity and variability, often linked to stronger blood flow. Conversely, if a feature falls significantly below the mean ($\mu - 0.5\sigma$), it suggests low variation and smoothness, potentially indicating poor blood flow due to low contrast or weak perfusion.

Figure 5 illustrates the dataset with quality labels, where Good quality signifies that the ROI has high quality, and Poor indicates weak quality. Throughout the rest of the work, good quality was numerically represented as 1, and poor quality as 0.

Table 1. Simulation result using testing data

Model	metric	Value in (%)
Support Vector Machine	accuracy	0.99 => (99%)
	precision	0.92 => (92%)
	recall	0.96 => (96%)
	F1-score	0.94 => (94%)
	false positive rate	0.007 => (0.7%)

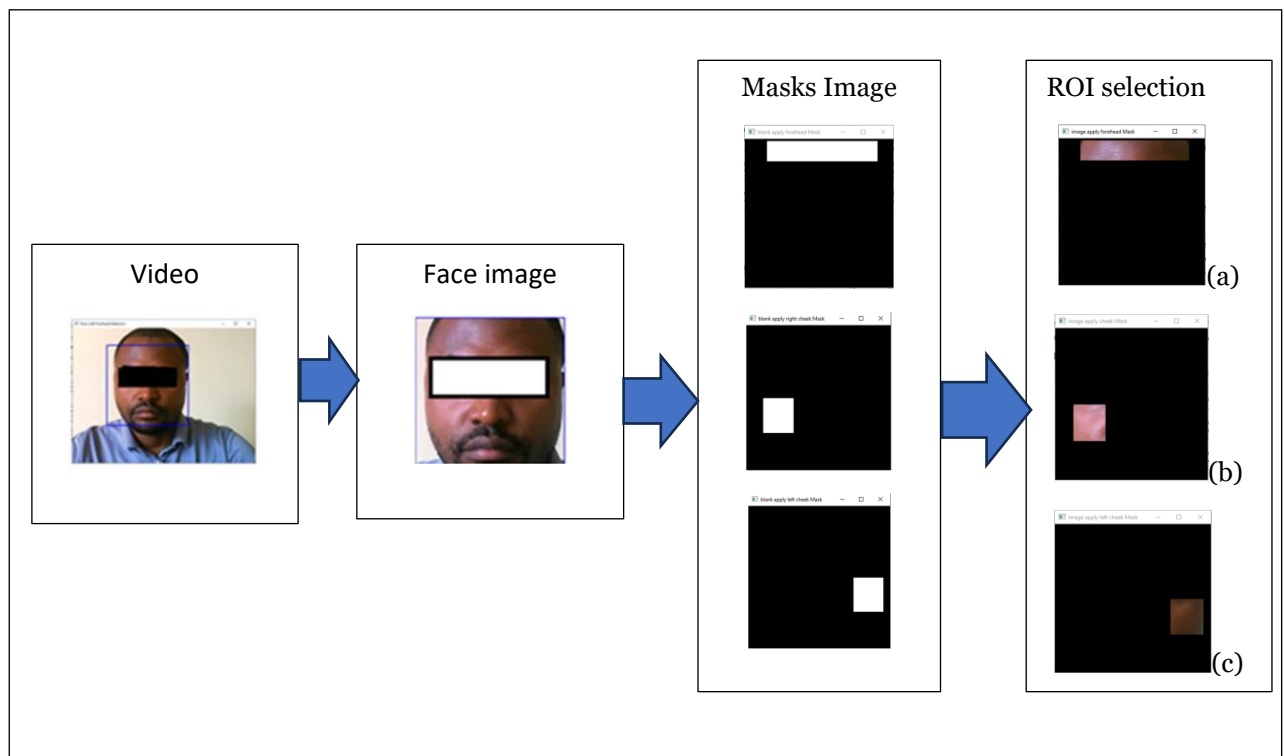


Figure 4. (a): forehead, (b): right cheek, (c): left cheek

Table 2. Characteristics for labelling the features

features	Quality	Characteristics of the image
Variance and standard deviation	High	Fluctuations in pixel intensity strong blood flow dynamics
	Low	Minimal fluctuation in pixel intensity weak blood flow
Average intensity	High	well-lit well-perfused skin regions
	Low	indicate poor lighting reduced blood perfusion
Signal-to-noise ratio	High	a strong physiological signal with minimal noise
	Low	Noisy or weak physiological signal
Entropy	High	rich texture active blood flow
	Low	Texture is smooth lacks complexity poor blood flow
Sobel filter	High	visible vascular structures
	Low	No strong vascular structures are visible
Gabor and Gaussian	High	Strong spatial frequency components in areas with good perfusion.
	Low	a lack of significant frequency

[6]:

	image name	ROI	mean Pixel intensity	signal to noise ratio	variance	standard deviation	sobel filter	gaussian filter	entropy	Gabor filter	quality_label
0	Image1392644182276178000.png	forehead	96.102708	2.668853	1296	36.008999	634.875255	97.290417	3.987436	1547.7368	1
1	Image1392644182276178000.png	left_cheek	59.245000	2.199917	725	26.930558	670.719671	60.057917	3.917957	956.7046	0
2	Image1392644182276178000.png	right_cheek	80.042500	2.469467	1050	32.412868	716.620399	81.202083	3.795118	1294.1350	0
3	Image1392644182309522000.png	forehead	96.617324	2.656602	1322	36.368758	650.077665	97.843568	3.958706	1556.7344	1
4	Image1392644182309522000.png	left_cheek	59.695784	2.199446	736	27.141276	688.710461	60.538896	3.897481	964.7088	0
...	...	...	...	...	...	...	...	...	...	...	...
2995	Image1392723949536884000.png	left_cheek	121.687028	2.156032	3185	56.440281	1319.243359	124.281400	4.191765	1984.0570	0
2996	Image1392723949536884000.png	right_cheek	104.487989	1.729320	3650	60.421444	1320.879848	106.045299	4.163277	1691.6222	0
2997	Image1392723949570229000.png	forehead	113.645848	1.752380	4205	64.852275	1148.542319	115.510295	4.778835	1840.2975	0
2998	Image1392723949570229000.png	left_cheek	121.799588	2.156258	3190	56.486546	1323.176315	124.403569	4.200033	1985.9454	0
2999	Image1392723949570229000.png	right_cheek	104.449554	1.723740	3671	60.594711	1323.273283	106.014413	4.159476	1691.0247	0

3000 rows × 11 columns

Figure 5. Dataset labelled with Good and Poor quality.

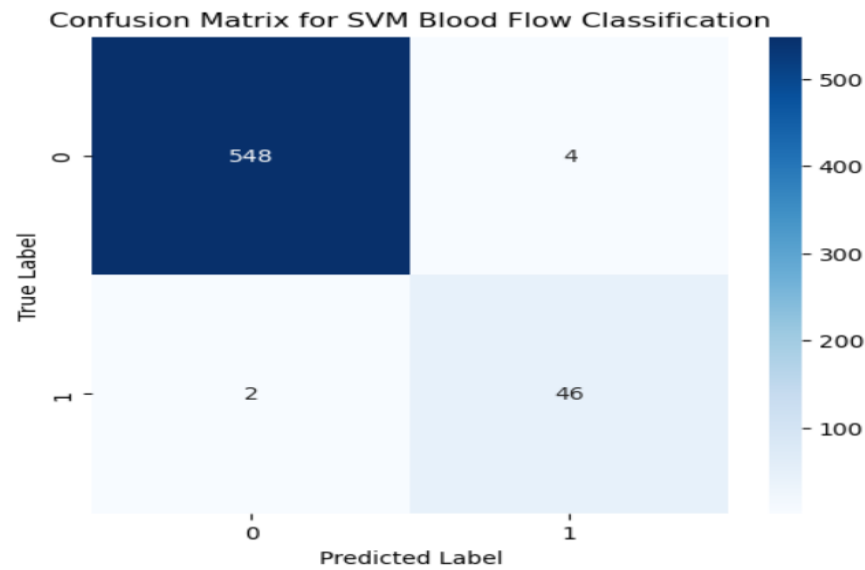


Figure 6. Confusion matrix for SVM blood flow quality classification from three ROIs on the face.

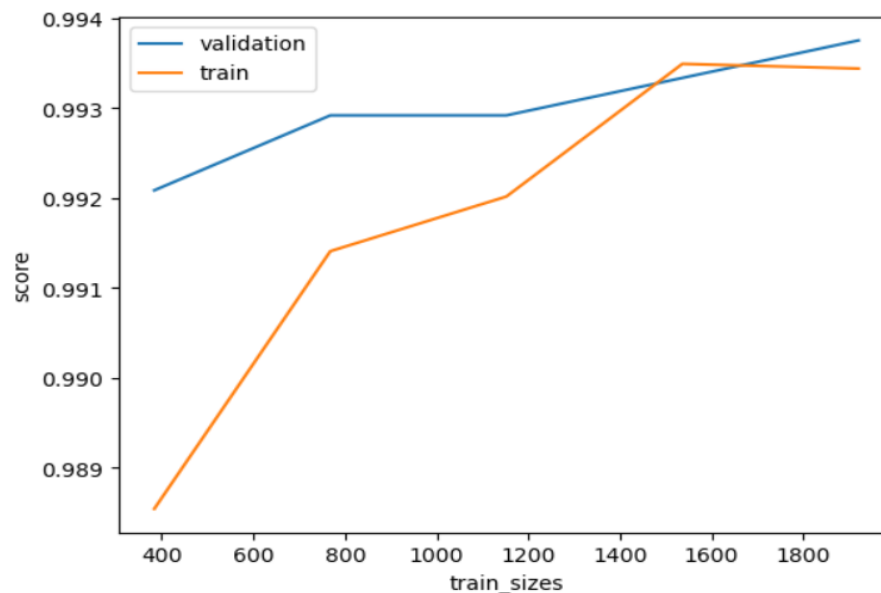


Figure 7. Training and validation accuracy of the model.

The model's outcomes are highly encouraging for our needs:

- Accuracy (99%): Exceptionally high, indicating the model's overall strong performance.
- Precision (92%): A high precision rate signifies that when the model predicts "Good" blood flow, it is accurate 92% of the time, which is crucial for reducing false positives.
- Recall (96%): A high recall rate shows the model's effectiveness in identifying actual instances of

"Good" blood flow, capturing 96% of them, which is excellent for ensuring no good samples are overlooked.

- F1-score (94%): This score reflects a well-balanced trade-off between precision and recall, indicating a strong balance between the two metrics.
- False Positive Rate (0.7%): Extremely low, suggesting the model seldom misclassifies "Poor" blood flow as "Good," thereby minimizing false alarms.

In the realm of real-time blood flow quality assessment, these metrics demonstrate the model's robustness. It efficiently identifies high-quality blood flow with minimal false positives and high recall, which is vital for precise monitoring in real-time applications.

5. Conclusion and future work

In this study, a computer vision model employing a machine learning approach for real-time analysis was introduced, allowing for the classification of a patient's blood flow quality in the regions of interest: the forehead, right cheek, and left cheek. The process begins by detecting the subject's face in the initial frame and maintaining its tracking throughout the video. Next, three mask regions are generated to facilitate feature extraction using the bitwise operation method. Finally, a Support Vector Machine model is utilized to determine which of the three areas exhibits good blood flow quality.

The proposed system shows promising results when evaluated using the PURE dataset. It delivered outstanding performance, achieving an accuracy rate of 99% and a False Positive Rate of merely 0.7%. This classification aims to identify the region of interest that contains the most vital physiological data. This will help us reduce errors in future work when estimating the rPPG signal.

In our upcoming research, we plan to utilize advanced artificial intelligence algorithms to predict the patient's rPPG signal, which will allow us to determine the heart rate.

References

- [1] F.-T.-Z. Khanam, A. Al-Naji, J. Chahl, Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: A review, *Appl. Sci.* 9 (20) (2019) 4474.
- [2] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, M. Fuchs, Assessment of deep learning based on blood pressure prediction from PPG and rPPG signals, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3820–3830.
- [3] C. Massaroni, A. Nicolo, M. Sacchetti, E. Schena, Contactless methods for measuring respiratory rate: A review, *IEEE Sens. J.* 21 (11) (2020) 12821–12839.
- [4] R. Yousefi, M. Nourani, Separating arterial and venous-related components of photoplethysmographic signals for accurate extraction of oxygen saturation and respiratory rate, *IEEE J. Biomed. Health Inf.* 19 (3) (2014) 848–857.
- [5] A. Gudi, M. Bittner, J. van Gemert, Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation, *Appl. Sci.* 10 (23) (2020) 8630.
- [6] L. Scalise, Non-contact heart monitoring, *Adv. Electrocardiogram. Methods Anal.* 4 (2012) 81–106.
- [7] A. Challoner, C. Ramsay, A photoelectric plethysmograph for the measurement of cutaneous blood flow, *Phys. Med. Biol.* 19 (3) (1974) 317.
- [8] L.A. Aarts, V. Jeanne, J.P. Cleary, C. Lieber, J.S. Nelson, S.B. Oetomo, W. Verkruysse, Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—A pilot study, *Early Hum. Dev.* 89 (12) (2013) 943–948.
- [9] A. Al-Naji, K. Gibson, S.-H. Lee, J. Chahl, Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods, *IEEE Access* 5 (2017) 15776–15790.
- [10] W. Verkruysse, L.O. Svaasand, J.S. Nelson, Remote plethysmographic imaging using ambient light, *Opt. Express* 16 (26) (2008) 21434–21445.
- [11] D.J. McDuff, J.R. Estepp, A.M. Piasecki, E.B. Blackford, A survey of remote optical photoplethysmographic imaging methods, in: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE*, 2015, pp. 6398–6404.
- [12] P.V. Rouast, M.T. Adam, R. Chiong, D. Cornforth, E. Lux, Remote heart rate measurement using low-cost RGB face video: a technical literature review, *Front. Comput. Sci.* 12 (2018) 858–872.
- [13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001, pp. I-511–I-518.
- [14] M.-Z. Poh, D. J. McDuff, and R.W. Picard, "Noncontact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Exp.*, vol.18, no.10,2010, Art. no.10762.
- [15] C. Tomasi and T. Kanade, "Shape and motion from image streams: A factorization method—Part 3

detection and tracking of point features,” Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech.Rep.CMU-CS-91 132, Apr. 1991.

- [16] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D.A. Clifton, and C. Pugh, “Non-contact video-based vital sign monitoring using ambient light and autoregressive models,” *Physiol. Meas.*, vol. 35, no. 5, pp.807–831,2014.
- [17] L. Iozzia, L. Cerina, and L. Mainardi, “Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA,” *Physiol. Meas.*, vol.37, no.11, pp. 1934–1944, 2016. C. Zhao, M. Zhou, W. Han, Y. Feng, Anti-motion remote measurement of heart rate based on region proposal generation and multi-scale ROI fusion, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–13.
- [18] J. Li, K. Vatanparvar, L. Zhu, J. Kuang, A. Gao, Enhancement of remote PPG and heart rate estimation with optimal signal quality index, in: 2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks, BSN, 2022, pp. 1–4.
- [20] W. Wei, K. Vatanparvar, L. Zhu, J. Kuang, A. Gao, Remote photoplethysmography and heart rate estimation by dynamic region of interest tracking, in: 2022 EMBC, 2022, pp. 3243–3248.
- [21] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M.G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, M. Grundmann, MediaPipe: A framework for building perception pipelines, 2019, arXiv:1906.08172.
- [22] T.P. Pagano, L.L. dos Santos, V.R. Santos, P.H.M. SÁ, Y.d.S. Bonfim, J.V.D. Paranhos, L.L. Ortega, L.F.S. Nascimento, A. Santos, M.M. Rönnau, I. Winkler, E.G.S. Nascimento, Remote heart rate prediction in virtual reality head-mounted displays using machine learning techniques, *Sensors* 22 (23) (2022).
- [23] K.L. Wong, J. Wei Chin, T.T. Chan, I. Odinaev, K. Suhartono, K. Tianqu, R.H.Y. So, Optimising rPPG signal extraction by exploiting facial surface orientation, in: 2022 CVPRW, 2022, pp. 2164–2170.
- [24] Ji-Hae Kim, Byung-Gyu Kim, Partha Pratim Roy, and Dami Jeong, “Efficient facial expression recognition algorithm based on hierarchical deep neural network structure,” *IEEE Access*, vol. 7, pp. 41273-41285, 2019.
- [25] Dami Jeong, Byung-Gyu Kim, and Suh-Yeon Dong, “Deep joint spatiotemporal network (DJSTN) for efficient facial expression recognition,” *Sensors*, vol. 20, no. 7, p. 1936, 2020, doi: 10.3390/s20071936.
- [26] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceeding of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001, doi: 10.1109/cvpr.2001.990517.
- [27] Stricker, R., Müller, S., Gross, H.-M. “Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot,” in: *Proc. 23rd IEEE Int. Symposium on Robot and Human Interactive Communication (Ro-Man 2014)*, Edinburgh, Scotland, UK, pp. 1056-1062, IEEE 2014.
- [28] Achancaray, D., et al., "Automatic Image Quality Assessment for Retinal Fundus Images Using Deep Learning," *Biomedical Signal Processing and Control*, vol. 76, p. 103671, 2022.
- [29] Y. Sun and N. Thakor, “Photoplethysmography revisited: From contact to noncontact, from point to imaging,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 463–477, Mar. 2016.
- [30] J. Allen, K. Howell, “Microvascular imaging: techniques and opportunities for clinical physiological measurements,” *Physiological measurement*, 35(7), R91. (2014).
- [31] S. G. Hoggar, *Mathematics of Digital Images: Creation, Compression, Restoration, Recognition*. Cambridge: Cambridge University Press, 2006.
- [32] K. Bredies and D. Lorenz, *Mathematical Image Processing*. Cham, Switzerland: Springer, 2019.
- [33] Gonzalez, R. C., and Woods, R. E., *Digital Image Processing* (4th Edition). Pearson, 2018.
- [34] "PURE dataset," [Online]. Available: <http://www.tu-ilmenau.de/neurob/data-sets/pulse>. [Accessed: Apr. 30, 2024].