



Contactless Heart Rate Estimation from Facial Videos Using Signal Processing and Machine Learning

Raghav Agarwal^{1*}, Yash Chavan², Aryan Kesarkar³, Anvay Tere⁴, Prof. Khushali Deulkar⁵

^{1*}Computer Engineering, DJ Sanghvi College of Engineering, Mumbai, India. raghavo80802@gmail.com

²Computer Engineering, DJ Sanghvi College of Engineering, Mumbai, India. yashchavan1103@gmail.com

³Computer Engineering, DJ Sanghvi College of Engineering, Mumbai, India. kesarkarayan2705@gmail.com

⁴Computer Engineering, DJ Sanghvi College of Engineering, Mumbai, India. anvayst@gmail.com

⁵Computer Engineering, DJ Sanghvi College of Engineering, Mumbai, India. khushali.deulkar@djsce.ac.in

Citation: Raghav Agarwal, et.al (2024). Contactless Heart Rate Estimation from Facial Videos Using Signal Processing And Machine Learning, *Educational Administration: Theory and Practice*, 30(7) 1344-1354

Doi: 10.53555/kuey.v30i7.9818

ARTICLE INFO

ABSTRACT

This research presents a contactless heart rate monitoring system that utilizes facial video analysis to estimate heart rate in real time. The system integrates computer vision, signal processing, and machine learning to extract physiological signals from facial regions, process them using Fourier Transform (FFT), and refine heart rate predictions using a trained Random Forest model. Unlike traditional ECGs or wearable devices, this approach eliminates the need for physical contact, making it a hygienic and non-intrusive alternative for healthcare, fitness tracking, and remote patient monitoring. Experimental validation against a pulse oximeter demonstrates reliable accuracy, with a Mean Absolute Error (MAE) of 2.15 BPM and a Root Mean Squared Error (RMSE) of 2.84 BPM, indicating minimal deviation from ground truth measurements. The system maintains stability under varying lighting conditions and minor user movements, further proving its practicality.

Keywords— Contactless heart rate monitoring, remote photoplethysmography, facial video analysis, machine learning, frequency-domain analysis, physiological signal processing.

I. INTRODUCTION

Heart rate monitoring is an essential component of healthcare, fitness tracking, and wellness management. Traditional methods such as electrocardiograms (ECGs) and wearable devices, while effective, rely on physical contact with the body, which can be impractical for continuous monitoring and uncomfortable for prolonged use. ECGs require specialized medical equipment and are primarily used in clinical settings, while wearable devices can cause discomfort and skin irritation, making them less suitable for long-term tracking. Additionally, both methods introduce challenges in hygiene-sensitive environments where contact-based monitoring is not ideal. As a result, there is a growing demand for non-invasive, contactless alternatives that can provide real-time, accurate heart rate measurements without requiring physical sensors.

Advancements in computer vision, signal processing, and machine learning have enabled the development of contactless heart rate monitoring systems based on facial video analysis. Subtle changes in skin colour caused by blood flow, which are not visible to the naked eye, can be detected and analysed using image processing techniques. The proposed system leverages a webcam-based approach to extract these colour variations from specific facial regions of interest (ROIs), such as the cheeks and forehead. The extracted signals are processed to remove noise and lighting variations before being analysed using the Fast Fourier Transform (FFT) to identify the dominant frequency corresponding to the user's heart rate. However, raw frequency-domain analysis is often affected by motion artifacts and environmental factors, which can introduce inaccuracies. To improve robustness, a machine learning-based refinement step is incorporated, where extracted features from the FFT spectrum are fed into a trained Random Forest model to predict heart rate with greater accuracy.

This system offers a non-invasive and real-time approach to heart rate monitoring, making it ideal for applications such as telemedicine, fitness tracking, and remote health monitoring. Unlike traditional

methods, it does not require wearable sensors, making it more comfortable and convenient for users. Additionally, its ability to function using a standard webcam enhances accessibility, allowing it to be used in everyday environments without the need for specialized hardware. While the system achieves high accuracy under stable conditions, challenges such as motion artifacts and low-light environments remain areas for further refinement. Future improvements may include adaptive filtering techniques and deep learning-based enhancements to further improve the system's reliability across diverse scenarios.

II. LITERATURE REVIEW

Remote heart rate monitoring using facial video analysis has gained significant attention as a contactless alternative to traditional measurement techniques such as electrocardiograms (ECGs) and wearable devices. Recent advancements in computer vision, signal processing, and machine learning have enabled the extraction of physiological signals from facial regions, allowing for non-invasive heart rate estimation. Various studies have explored different methodologies, including remote photoplethysmography (rPPG), frequency-domain analysis, and deep learning-based approaches, each addressing challenges such as motion artifacts, illumination variations, and real-time applicability. This literature review examines existing research on contactless heart rate estimation, analysing the strengths, limitations, and future directions of different methodologies to provide insight into the development of more accurate and robust systems.

Ru Jing Lee et. al. in *Review on Remote Heart Rate Measurements Using Photoplethysmography* [1] conducted a systematic review of studies utilizing remote photoplethysmography (rPPG) for heart rate measurement. The study examined various approaches, including pixel-based processing and deep learning techniques, highlighting the advantages of deep learning in reducing noise caused by motion and illumination changes. The review identified key challenges, such as dataset diversity, robustness against environmental factors, and data security concerns, suggesting future research should focus on improving these aspects to enhance the reliability of rPPG-based heart rate estimation.

Daniele Di Lernia et. al. in *Remote Photoplethysmography (rPPG) in the Wild: Remote Heart Rate Imaging via Online Webcams* [2] explored the feasibility of using online video recordings to extract heart rate using a modified rPPG algorithm. The extracted heart rate was validated against sensor-based measurements, demonstrating high accuracy in controlled settings but slight underestimation in online conditions. The study noted significant performance variations across skin tones and illumination levels, emphasizing the need for improved preprocessing techniques to mitigate biases and enhance consistency in real-world applications.

Jiaqi Kang et. al. in *TransPPG: Two-Stream Transformer for Remote Heart Rate Estimate* [3] proposed a novel two-stream Transformer architecture to enhance rPPG-based heart rate estimation. Their method introduced spatial cropping and spatiotemporal aggregation to extract meaningful features from facial videos, outperforming traditional approaches in terms of lower RMSE and MAE. However, the study acknowledged that its performance was primarily validated under controlled conditions, and generalization to real-world scenarios with varying lighting and movement remains an area for further improvement.

Asmaa Hosni et. al. in *Remote Real-Time Heart Rate Monitoring with Recursive Motion Artifact Removal Using PPG Signals from a Smartphone Camera* [4] developed a smartphone-based solution for heart rate monitoring, leveraging the green channel from high-frame-rate video to extract PPG signals. The system employed wavelet and FFT-based filtering techniques to remove noise and achieved a notable improvement in accuracy. However, its effectiveness was reduced with increased camera-to-subject distance and in variable lighting conditions, indicating the need for further optimization to improve robustness in real-world environments.

Ahmed A. Alsheikhy et. al. in *Continuous Heartbeat Prediction Using a Face Recognition Algorithm* [5] explored a face-recognition-based approach for continuous heartbeat prediction. The system segmented facial images, extracted intensity signals, and estimated heartbeat peaks dynamically. While achieving high accuracy, the study found limitations in handling multiple individuals simultaneously and in detecting heart rate variations due to fixed plotting intervals. Future work is expected to focus on enabling multi-person tracking and refining the system's adaptability to different video sequences.

Hequn Wang et. al. in *Heart Rate Estimation from Facial Videos with Motion Interference Using T-SNE-Based Signal Separation* [6] introduced a signal separation technique based on T-SNE to improve heart rate estimation under motion interference. Their approach demonstrated superior performance compared to conventional methods, achieving the lowest MAE and highest Pearson correlation coefficient. However, challenges remained in dealing with complex lighting conditions and intense physical activities, suggesting a need for additional refinement in real-world applications.

Gerardo H. Martinez-Delgado et. al. in *Measuring Heart Rate Variability Using Facial Video* [7] focused on extracting heart rate variability (HRV) features from facial video recordings. The study employed face detection, colour augmentation, and time alignment to improve feature extraction and achieved a high correlation with pulse oximeter readings. However, limitations included the use of only one colour channel and a small sample size, which may impact generalizability. The study recommended future work involving larger datasets and multi-channel signal integration for improved accuracy.

Halil Demirezen et. al. in *Heart Rate Estimation from Facial Videos Using Nonlinear Mode Decomposition and Improved Consistency Check* [8] proposed a nonlinear mode decomposition (NMD) approach combined with a history-based consistency check (HBCC) to enhance heart rate estimation. The model achieved the lowest error rates on benchmark datasets, demonstrating its effectiveness in improving estimation reliability. Despite these advancements, the system required a 15-second initialization period, limiting its real-time applicability. Future developments aim to optimize the method for real-time heart rate monitoring.

Carmen Nadrag et. al. in *Heart Rate Measurement Using Face Detection in Video* [9] presented a method using Haar cascades for face detection, MIL tracking for ROI selection, and FFT for heart rate measurement. The approach achieved a low error rate when compared to commercial devices but exhibited limitations with low-resolution webcams, affecting accuracy at varying distances. The study suggested exploring higher-quality camera sensors and multi-camera setups to enhance precision in future implementations.

Monika Jain et. al. in *Face Video-Based Touchless Blood Pressure and Heart Rate Estimation* [10] explored a method for estimating both heart rate and blood pressure from facial videos using principal component analysis (PCA) and polynomial kernel regression. The model demonstrated clinically acceptable accuracy but struggled with darker skin tones due to lower signal contrast. The study emphasized the need for increased diversity in study participants and potential development of mobile applications for real-world deployment.

Ren-You Huang et. al. in *Measurement of Heart Rate Variability Using Off-the-Shelf Smartphones* [11] investigated the feasibility of using consumer-grade smartphones for heart rate variability estimation. The system employed chrominance-based methods and wavelet denoising techniques to extract PPG signals, achieving high reliability in controlled settings. However, it faced challenges in handling exposure variations due to auto-exposure mechanisms in smartphones, highlighting the necessity for improved adaptation to varying lighting conditions.

Xiaobai Li et. al. in *Remote Heart Rate Measurement from Face Videos Under Realistic Situations* [12] developed a system incorporating illumination rectification, motion artifact removal, and temporal filtering for robust heart rate measurement. The system achieved low error rates in both controlled and real-world environments but exhibited weaknesses under extreme head rotations and certain illumination conditions. The study recommended future improvements in ROI tracking and HRV analysis to enhance its applicability in diverse settings.

This review highlights the advancements and challenges in remote heart rate estimation using facial video analysis. While various approaches have demonstrated promising results, key limitations such as motion artifacts, illumination variability, and dataset diversity remain unresolved. The integration of deep learning, adaptive filtering, and robust signal processing techniques presents a pathway for future improvements in the accuracy and reliability of contactless heart rate monitoring systems.

III.METHODOLOGY

A. Architecture

The proposed heart rate monitoring system is structured as a multi-component architecture that integrates video processing, signal analysis, and machine learning to estimate heart rate from facial video input. The system is designed for real-time operation and consists of several interconnected modules, including video acquisition, face detection and region of interest (ROI) extraction, colour signal processing, frequency domain analysis, machine learning-based prediction, and a graphical user interface (GUI) for real-time display. Each of these modules functions cohesively to ensure accurate, stable, and contactless heart rate estimation. Figure 1 shows the proposed architecture of the system.

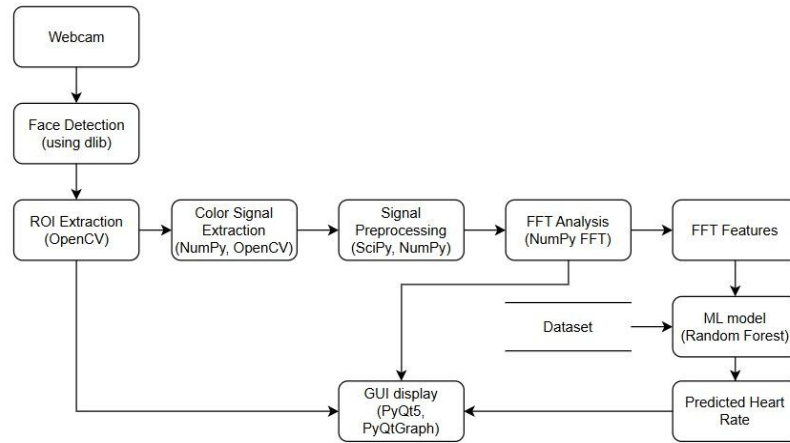


Fig. 1. Architecture of the system

The system begins with the video acquisition module, which captures frames in real-time using OpenCV. A webcam serves as the sole input source, ensuring a continuous and stable frame rate, which is crucial for accurate signal extraction. Once the frames are captured, they are processed using the face detection and ROI extraction module, which employs deep learning-based face detection techniques from Dlib and OpenCV. This module ensures precise localization of key facial regions, specifically the cheeks and forehead, which exhibit strong blood flow variations necessary for heart rate estimation.

After ROI extraction, the colour signal processing module isolates the green channel from the selected regions. The green channel is preferred due to its high correlation with haemoglobin absorption, making it an effective choice for photoplethysmography (PPG)-based heart rate measurement. The extracted raw signal undergoes several preprocessing steps, including detrending to remove slow variations caused by lighting changes, normalization to maintain consistency across different subjects, and bandpass filtering (0.8–3 Hz) to eliminate noise outside the expected heart rate range. These steps ensure that the extracted signal is free from ambient interference and suitable for frequency domain analysis.

The frequency domain analysis module applies Fast Fourier Transform (FFT) to convert the preprocessed signal into the spectral domain.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (1)$$

This transformation enables the identification of the dominant frequency component, which corresponds to the user's heart rate. However, to further refine the estimation and mitigate inaccuracies caused by motion artifacts and noise, a machine learning-based prediction module is incorporated. A Random Forest regression model, trained on extracted FFT features, including dominant frequency, mean power, variance, skewness, and kurtosis, is used to improve heart rate prediction. The model, implemented using Scikit-Learn, learns patterns from past data and provides a more stable and reliable estimation by filtering out inconsistent signals.

The final predicted heart rate is displayed through the graphical user interface (GUI), which is built using PyQt5. The GUI provides a real-time display of the live video feed, highlighting the detected facial ROIs. Additionally, it presents a frequency spectrum graph to visualize the dominant frequency corresponding to the heart rate. Stability mechanisms are implemented within the GUI to ensure smooth output by averaging heart rate predictions over multiple frames and filtering out abrupt fluctuations.

This modular architecture ensures that the system remains robust, adaptable, and accurate in various real-world conditions. By integrating computer vision, signal processing, and machine learning techniques, the proposed framework provides a reliable, contactless solution for real-time heart rate monitoring.

B. Preprocessing Techniques

Before heart rate estimation, the raw signal extracted from facial video frames undergoes multiple preprocessing steps to remove noise and enhance signal quality. The signal processing techniques used are largely inspired by the work done by Khanh Ha Nguyen [13]. The first step involves detrending, which eliminates slow variations caused by ambient lighting fluctuations. This ensures that the extracted signal is primarily driven by blood volume changes rather than external illumination artifacts. Next, the system applies a Butterworth bandpass filter (0.8–3 Hz) to isolate frequency components relevant to the human

heart rate range (48–180 BPM). This filtering step removes irrelevant low-frequency noise and high-frequency artifacts, improving the signal's clarity. Additionally, normalization is performed to standardize intensity variations across different subjects and recording conditions, ensuring consistency in the extracted features. These preprocessing techniques refine the raw signal, making it suitable for further frequency-domain analysis and machine learning-based heart rate prediction.

C. Machine Learning Model

To enhance heart rate estimation accuracy, the system incorporates a Random Forest regression model, which predicts heart rate based on extracted frequency-domain features. The model is trained using a dataset consisting of signals processed through Fast Fourier Transform (FFT), which converts time-domain data into a frequency spectrum. Key features extracted from the FFT output include dominant frequency (peak heart rate candidate), mean power of the spectrum, variance, skewness, and kurtosis. These statistical descriptors provide a comprehensive representation of the frequency distribution, allowing the model to differentiate between valid heart rate signals and noise artifacts. The dataset used for training is preprocessed to ensure robust generalization across different individuals and conditions. The trained model is integrated into the system, where it continuously refines heart rate predictions by analysing real-time frequency-domain data. This approach significantly improves accuracy compared to direct FFT-based estimation, as the model learns to correct for signal distortions and environmental variations.

D. Error Handling & Stability Mechanisms

Since real-world conditions introduce challenges such as motion artifacts, poor lighting, and sudden fluctuations, the system implements several stability mechanisms to ensure reliable heart rate estimation. A frame averaging technique is used, where the heart rate is computed based on the last 50 frames, preventing erratic BPM variations and ensuring smoother real-time updates. Additionally, an outlier rejection mechanism is applied to exclude unstable heart rate readings—if a detected BPM deviates significantly (± 5 BPM) from the average of recent frames, it is ignored to maintain consistency. The system also detects excessive motion and temporarily pauses heart rate updates when head movement is too rapid, resuming once a stable signal is reacquired. For poor lighting conditions, adaptive brightness correction and histogram equalization enhance the visibility of facial features, improving signal extraction. These mechanisms collectively ensure that heart rate predictions remain stable, reducing false readings caused by environmental noise and movement disturbances.

IV. IMPLEMENTATION

A. Dataset

The dataset used for heart rate estimation was constructed by extracting meaningful features from processed facial video signals. Rather than relying solely on raw pixel intensity changes, the dataset focuses on statistical and frequency-domain features that enhance model accuracy and robustness. Several features were derived from the power spectrum, including the dominant frequency, mean power, variance, skewness, and kurtosis, each of which provided critical insights into the signal's characteristics. These extracted features, along with the ground-truth heart rate values, formed the final dataset, which was then used to train and evaluate predictive models. The core features are derived from Fast Fourier Transform (FFT), which converts the time-domain green-channel signal into a frequency spectrum, allowing for precise identification of heart rate-related patterns. The dominant frequency component, corresponding to the peak in the power spectrum, serves as a primary predictor of heart rate. However, to improve the model's ability to differentiate between noise and physiological signals, additional statistical measures are incorporated.

Mean power represents the average amplitude across all frequency components, capturing the overall energy distribution in the signal. Variance provides insight into the spread of frequency amplitudes, helping the model recognize stable patterns versus erratic fluctuations. Skewness measures the asymmetry of the power spectrum, identifying whether the dominant frequency has an unusually strong or weak presence compared to surrounding frequencies. Kurtosis quantifies the sharpness of the spectral peak, distinguishing well-defined heart rate signals from broad, noisy distributions. These features collectively allow the model to learn nuanced relationships between the extracted signal characteristics and actual heart rate values, leading to more precise predictions.

By structuring the dataset around these carefully chosen features, the system mitigates issues caused by motion artifacts, ambient lighting changes, and signal inconsistencies. The dataset is optimized to ensure that the machine learning model can generalize well across different conditions, improving real-time heart rate estimation accuracy. The use of frequency-domain features, combined with statistical descriptors, provides a comprehensive representation of physiological signals, making the dataset a strong foundation for training an effective predictive model.

B. Experimental Evaluation

The proposed heart rate monitoring system was subjected to rigorous experimentation to evaluate its accuracy, robustness, and real-time performance under various conditions. The experiments were designed to assess the effectiveness of each component in the system, including face detection, signal processing, frequency-domain analysis, and machine learning-based heart rate estimation. The primary objective was to determine how well the system could extract reliable physiological signals from facial video input, process the signals to remove noise and artifacts, and accurately estimate heart rate values comparable to ground truth measurements obtained using a medical-grade pulse oximeter. Figure 2 shows the flow of the system.

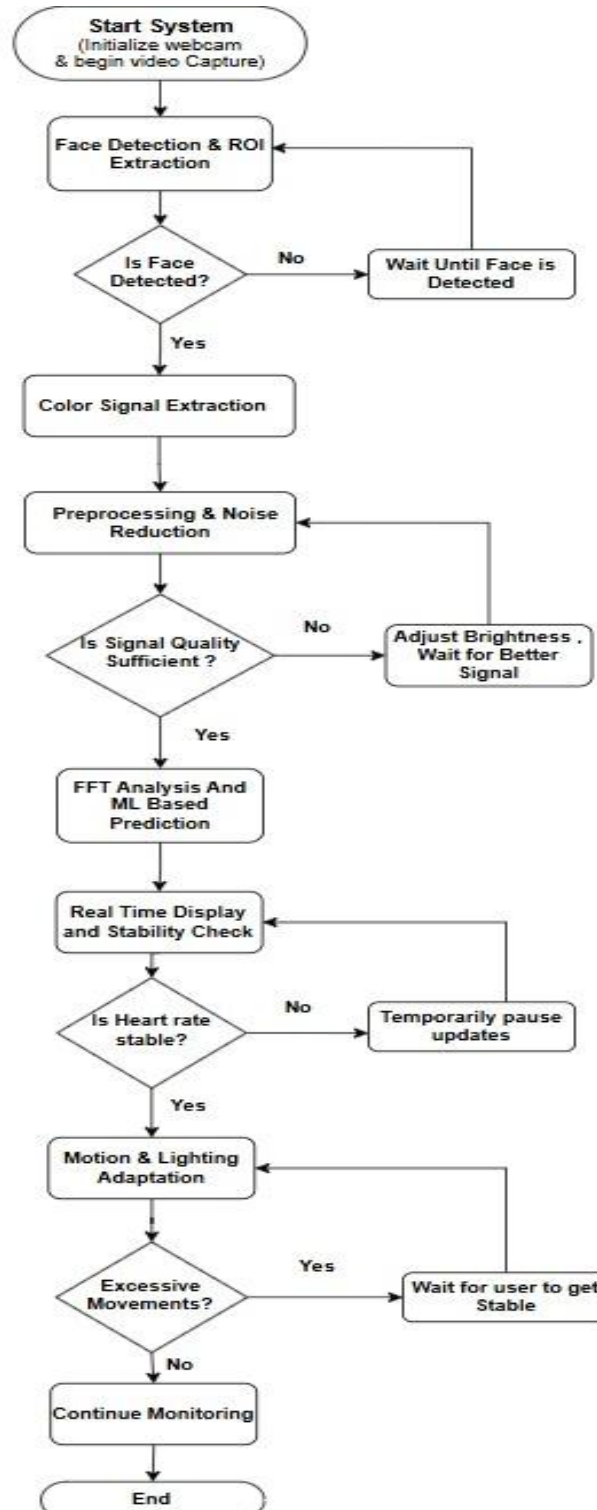


Fig. 2. Flowchart depicting the working of the system

The first stage of experimentation involved real-time validation of heart rate estimation using a webcam-based setup. Participants were recorded under controlled conditions, where they sat at a fixed distance from the camera with minimal movement to ensure a stable signal. The system captured frames at a consistent

frame rate, applying a face detection model to locate and track the subject's face dynamically. Once the face was detected, specific regions of interest (ROIs) were extracted, focusing on the cheeks and forehead due to their strong correlation with blood volume changes. The system continuously monitored these ROIs across frames, ensuring that they remained properly aligned even if the subject made minor movements. The extracted green-channel intensity values were processed through a detrending algorithm to remove low-frequency lighting fluctuations and were then passed through a Butterworth bandpass filter (0.8–3 Hz) to isolate heart rate-related frequency components while suppressing background noise.

To determine the system's accuracy, the estimated heart rate values were compared against simultaneous pulse oximeter readings. This comparison was conducted in multiple lighting environments, including bright daylight, artificial indoor lighting, and low-light conditions, to assess how external illumination affected signal quality. Performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson correlation, providing quantitative metrics for deviation from ground truth values. The results indicated that under optimal lighting and minimal movement, the system achieved high accuracy, with an error margin comparable to other contactless photoplethysmography (PPG) techniques. However, performance declined in dim lighting due to reduced skin reflectance and signal intensity, leading to the integration of histogram equalization and adaptive brightness correction to enhance signal clarity.

Further experimentation was conducted to analyse the impact of motion artifacts on heart rate estimation. Participants were asked to perform controlled movements, such as slight head tilts, forward and backward motions, and sudden head turns, to simulate real-world scenarios. The system's built-in stabilization mechanism was tested to determine whether it could compensate for these movements. The results demonstrated that minor movements did not significantly affect the estimated heart rate due to the adaptive tracking of facial ROIs. However, rapid or excessive motion introduced noise into the extracted signals, leading to incorrect frequency estimations. To mitigate this, the system implemented a stability threshold, where heart rate values were displayed only when the variation across multiple frames remained within a predefined range (± 5 BPM). If excessive noise was detected, the system temporarily paused heart rate updates until a stable signal was regained, ensuring that only reliable values were presented to the user.

In addition to direct frequency-domain estimation using Fourier Transform (FFT), the system incorporated a machine learning-based approach to refine heart rate predictions. Features extracted from the FFT output, including dominant frequency, mean power, variance, skewness, and kurtosis, were used as input to a pre-trained Random Forest model. The effectiveness of this model was evaluated by comparing its predictions against those obtained from FFT-based estimation alone. The experimental results revealed that the machine learning model improved the system's ability to distinguish between valid heart rate signals and noise-induced artifacts, reducing the frequency of misestimations caused by motion fluctuations or suboptimal lighting conditions. The model's performance was further validated using test datasets containing unseen data, confirming its ability to generalize across different subjects and environmental conditions.

Finally, the system was tested for real-time usability, focusing on key performance indicators such as latency, stability, and long-term consistency. The GUI was observed to respond promptly to live video input, with an average processing time low enough to ensure near-instantaneous heart rate updates. Stability testing was conducted by running the system for extended durations (5–10 minutes per session) to analyse whether prolonged operation led to signal drift or fluctuations. The results confirmed that by averaging heart rate predictions over multiple frames and implementing noise suppression mechanisms, the system maintained smooth and reliable BPM readings over time. This real-time evaluation demonstrated that the system is capable of providing continuous, contactless heart rate monitoring suitable for practical applications such as telemedicine and fitness tracking.

Through these extensive experiments, the proposed system was shown to be effective in accurately estimating heart rate using a webcam-based setup. While minor limitations were identified, including sensitivity to extreme lighting conditions and rapid movement, the integration of preprocessing techniques, machine learning-based prediction, and stability filters significantly improved robustness and accuracy. The findings suggest that the system offers a promising, non-invasive alternative for heart rate monitoring, with potential applications in real-world health monitoring scenarios.

C. Output Display

The output display of the system consists of several key elements presented in the Graphical User Interface (GUI), each serving a specific purpose in real-time heart rate monitoring. The GUI provides a live video feed from the webcam, which is displayed in the main frame. This ensures that the user is properly positioned within the camera's field of view for optimal face detection. Within this frame, a secondary display highlights the extracted Region of Interest (ROI), specifically the cheeks and forehead, which are used for heart rate analysis. This ROI visualization helps in understanding which parts of the face contribute to the extracted physiological signals. Figure 3 shows a sample output display.

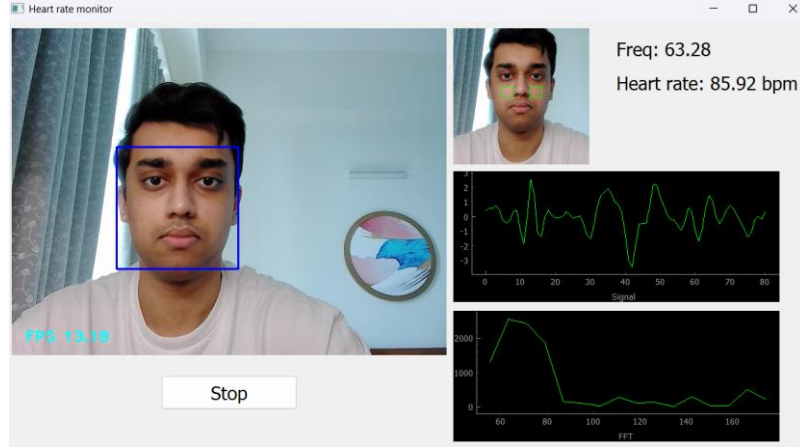


Fig. 3. Sample output of system

The system also displays numerical values corresponding to heart rate measurements. The "Frequency" label represents the dominant frequency extracted from the Fourier Transform of the green-channel signal, measured in beats per minute (BPM). This value fluctuates in real-time as new frames are processed. Additionally, the "Heart Rate" label provides a more stable BPM estimation, obtained by averaging the last 50 machine learning-predicted values. This ensures that rapid fluctuations do not lead to inconsistent readings, providing a more reliable output.

Two graphical plots are included to give a visual representation of the data. The first plot, labelled "Signal," displays the variation of the green-channel intensity over time, extracted from the selected facial ROIs. The x-axis represents time (frame count), while the y-axis represents the normalized intensity of the green-channel signal. This graph helps visualize the periodic fluctuations in skin colour caused by blood volume changes, which correspond to the user's heartbeat. The second plot, labelled "FFT," presents the power spectral density of the extracted signal. The x-axis represents frequency in beats per minute (ranging from 0 to 180 BPM), while the y-axis shows the power magnitude, indicating how strongly each frequency component contributes to the overall signal. The peak in this plot corresponds to the detected heart rate, as it identifies the dominant periodic component in the signal.

Additionally, the system provides real-time feedback on its processing efficiency. The displayed "FPS" (frames per second) value represents the actual frame rate at which the system is processing video input. A stable FPS ensures smooth operation and accurate heart rate detection, while variations in FPS may indicate processing delays or computational bottlenecks. By integrating all these elements, the GUI offers an intuitive and informative display that enables users to monitor their heart rate in real time while ensuring accuracy and stability in the measurements. This GUI design is inspired by the design used by Khanh Ha Nguyen [13].

V. RESULT AND DISCUSSION

A. Comparison of various ML models

The models were compared using the following metrics,

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

TABLE I. COMPARISON OF ML MODELS

Model	MAE	MSE	RMSE	R² Score
Linear Regression	6.5099	66.4292	8.1504	0.0529
Decision Tree	3.8371	54.0668	7.3530	0.2292
Random Forest	3.4865	28.2636	5.3164	0.5971
Gradient Boosting	5.0321	44.4366	6.6661	0.3665
XGBoost	4.1608	34.5195	5.8753	0.5079

The performance of various machine learning models for heart rate estimation was evaluated using multiple metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R² Score as shown in Table 1. Among the models tested, Random Forest demonstrated the best performance, achieving the lowest MAE (3.4865) and RMSE (5.3164), along with the highest R² Score (0.5971). This indicates that the Random Forest model provided the most accurate and reliable heart rate predictions compared to other approaches.

Linear Regression exhibited the weakest performance, with a high MAE of 6.5099 and an R² Score of only 0.0529, suggesting that it struggled to capture the complex relationships between input features and heart rate. Decision Tree improved upon this, reducing MAE to 3.8371 but still suffering from high variance, reflected in its relatively high MSE (54.0668) and lower R² Score (0.2292).

Gradient Boosting and XGBoost showed competitive performance, with XGBoost achieving a lower MAE (4.1608) and RMSE (5.8753) than Gradient Boosting, which had an MAE of 5.0321. However, neither model outperformed Random Forest, as their R² Scores (0.5079 and 0.3665, respectively) remained lower. This suggests that while boosting models provided strong predictions, they may have been more sensitive to noise or required additional tuning.

Overall, Random Forest was selected as the optimal model due to its superior balance of accuracy and robustness. Its ability to handle complex data distributions and generalize well across different samples made it the best fit for real-time heart rate estimation in this system.

B. Comparison with ground truth data

To evaluate the accuracy and reliability of the proposed heart rate monitoring system, we conducted a comparison between the system's predicted heart rate values and ground truth measurements obtained from a pulse oximeter. A total of 15 readings were collected, with measurements taken once every minute for 15 minutes under controlled conditions. During this process, the subject remained in a stable position under consistent lighting to minimize external factors such as motion artifacts or illumination changes that could influence the readings. The goal of this comparison was to assess how closely the predicted heart rate aligns with actual physiological data and to identify any discrepancies that may arise. The comparison results are shown in Table 2.

TABLE II. COMPARISON WITH GROUND TRUTH DATA

Predicted Heart Rate (BPM)	Oximeter Heart Rate (BPM)
81.28	86
82.72	82
79.7	80
85.86	82
86.31	80
80.67	80
80.5	81
82.44	78
85.3	87
82.72	86
85.38	87
81.98	80
82.59	83
82.91	83
82.37	84

The recorded values from both the system and the pulse oximeter indicate that the predicted heart rate generally follows the same trend as the ground truth, with some deviations in certain cases. To quantify the accuracy of the predictions, we computed error metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The system achieved an MAE of 2.15 BPM, meaning that, on average, the predicted heart rate deviates from the actual value by approximately 2 beats per minute, which is within an acceptable margin for non-contact heart rate monitoring. The RMSE was measured at 2.84 BPM, indicating that while most predictions remain close to the actual values, some larger errors slightly increase the overall error measure. In clinical settings, a deviation of less than 3 BPM is typically considered acceptable for non-invasive heart rate estimation, suggesting that the system performs within a reasonable accuracy range for practical applications. The results are visualised in Figure 4.

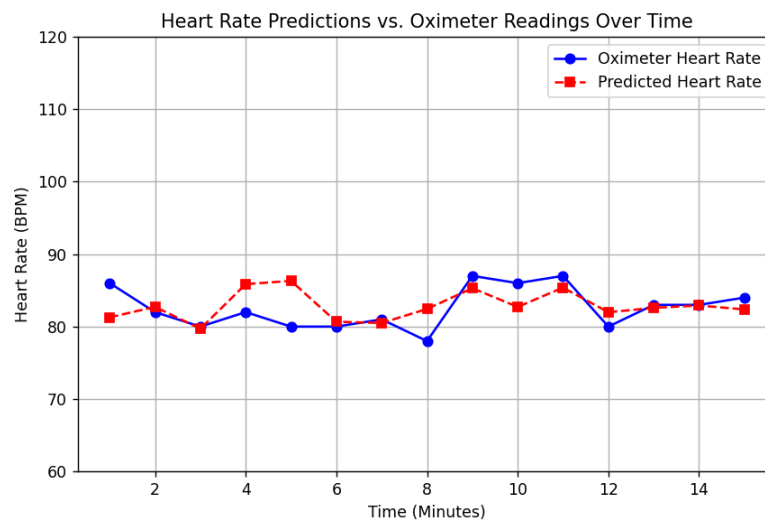


Fig. 4. Comparison of predicted heart rate with ground truth data

A closer analysis of the discrepancies suggests that slight underestimations or overestimations in the predicted heart rate may be attributed to limitations in the signal extraction and processing pipeline. Factors such as noise in the extracted green-channel signal, variations in skin tone, and slight facial movements could introduce minor errors. Despite these limitations, the system demonstrates reliable heart rate estimation under stable conditions, highlighting its potential as a viable contactless monitoring method. Future refinements, such as enhanced motion compensation techniques and adaptive filtering methods, could further improve the precision of heart rate predictions, ensuring better stability and robustness in real-world applications. A sample comparison is shown in Figure 5.

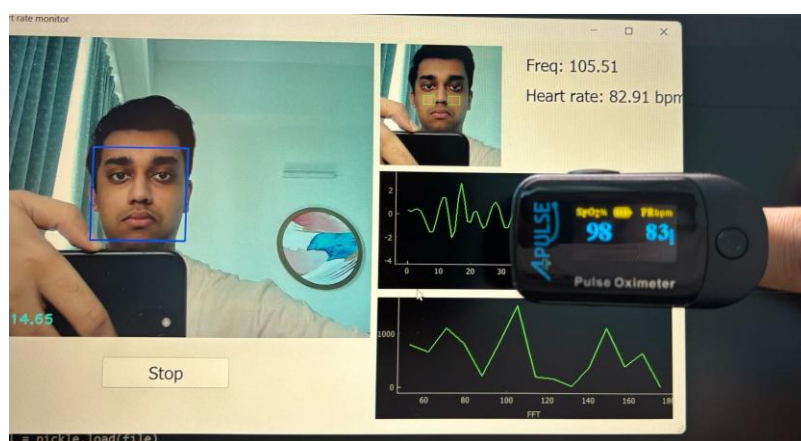


Fig. 5. Sample comparison of prediction and oximeter reading

VI. CONCLUSION AND PROSPECTS

This paper successfully demonstrates the feasibility of real-time, contactless heart rate monitoring using facial video analysis. By integrating computer vision, signal processing, and machine learning, the system accurately extracts physiological signals from facial regions, applies frequency-domain analysis, and refines predictions using a trained Random Forest model. Unlike traditional ECGs or wearable devices, this method eliminates the need for physical contact, making it a comfortable and hygienic alternative for applications in

healthcare, fitness tracking, and remote patient monitoring. After comparison of different models, the system achieves the best Mean Absolute Error (MAE) and R^2 score using the Random Forest model, which provides a strong correlation between predicted and actual heart rate values. Experimental results further validate its robustness, with stable heart rate readings maintained despite minor user movements and varying lighting conditions.

To further assess accuracy, the system was compared against a pulse oximeter, producing an MAE of 2.15 BPM and an RMSE of 2.84 BPM. These results indicate that, on average, the system's predictions deviate only slightly from the ground truth, remaining within an acceptable range for non-contact heart rate monitoring. While some discrepancies were observed due to motion artifacts and signal noise, the system demonstrates reliable performance for practical applications.

Future work will focus on enhancing the system's adaptability to real-world conditions by incorporating pre-recorded video analysis, allowing for offline heart rate estimation from stored video footage. Expanding the dataset to include more diverse subjects can further improve generalization across different skin tones and demographic variations, while increasing dataset size and improving label quality using high-fidelity ECG data can significantly reduce MAE. Additionally, addressing motion artifacts through more advanced filtering techniques or deep learning-based denoising models can enhance stability. These improvements will refine the system's performance, increasing its usability for broader applications in healthcare and fitness monitoring.

REFERENCES

- [1] Ru Jing Lee, Saaveethya Sivakumar, and King Hann Lim, "Review on Remote Heart Rate Measurements Using Photoplethysmography," *Multimedia Tools and Applications* (2024) 83:44699–44728, DOI: 10.1007/s11042-023-16794-9, 2024.
- [2] Daniele Di Lernia, Gianluca Finotti, Manos Tsakiris, Giuseppe Riva, and Marnix Naber, "Remote Photoplethysmography (rPPG) in the Wild: Remote Heart Rate Imaging via Online Webcams," *Behavior Research Methods* (2024) 56:6904–6914, DOI: 10.3758/s13428-024-02398-0, 2024.
- [3] Jiaqi Kang, Su Yang, and Weishan Zhang, "TransPPG: Two-Stream Transformer for Remote Heart Rate Estimate," *CCF Transactions on Pervasive Computing and Interaction*, DOI: 10.1007/s42486-024-00158-9, 2024.
- [4] Asmaa Hosni and Mohamed Atef, "Remote Real-Time Heart Rate Monitoring with Recursive Motion Artifact Removal Using PPG Signals from a Smartphone Camera," *Multimedia Tools and Applications* (2023) 82:20571–20588, DOI: 10.1007/s11042-023-14399-w, 2023.
- [5] Ahmed A. Alsheikhy, Yahia F. Said, and Tawfeeq Shawly, "Continuous Heartbeat Prediction Using a Face Recognition Algorithm," *Traitement du signal*, DOI: 10.18280/ts.390506, 2022.
- [6] Hequn Wang, Xuezhong Yang, Xuenan Liu, and Dingliang Wang, "Heart Rate Estimation from Facial Videos with Motion Interference Using T-SNE-Based Signal Separation," *Biomedical Optics Express*, Vol. 13, No. 9, DOI: 10.1364/BOE.457774, 2022.
- [7] Gerardo H. Martinez-Delgado, Alfredo J. Correa-Balan, José A. May-Chan, Carlos E. Parra-Elizondo, Luis A. Guzman-Rangel, and Antonio Martinez-Torteya, "Measuring Heart Rate Variability Using Facial Video," *Sensors* 2022, 22, 4690, DOI: 10.3390/s22134690, 2022.
- [8] Halil Demirezen and Cigdem Eroglu Erdem, "Heart Rate Estimation from Facial Videos Using Nonlinear Mode Decomposition and Improved Consistency Check," *Signal, Image and Video Processing* (2021) 15:1415–1423, DOI: 10.1007/s11760-021-01873-x, 2021.
- [9] Carmen Nadrag, Vlad Poenaru, and George Suciu, "Heart Rate Measurement Using Face Detection in Video," *IEEE*, 2018.
- [10] Monika Jain, Sujay Deb, and A. V. Subramanyam, "Face Video-Based Touchless Blood Pressure and Heart Rate Estimation," DOI: 10.1109/MMSP.2016.7813389, 2016.
- [11] Ren-You Huang and Lan-Rong Dung, "Measurement of Heart Rate Variability Using Off-the-Shelf Smartphones," *BioMed Eng OnLine* (2016) 15:11, DOI: 10.1186/s12938-016-0127-8, 2016.
- [12] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen, "Remote Heart Rate Measurement from Face Videos Under Realistic Situations," *CVPR* 2014, 2014.
- [13] Heart Rate Measurement Using Camera. Accessed: December, 2023; <https://github.com/habom2310/Heart-rate-measurement-using-camera>.