Enhancing Heart Rate Prediction through Deep CNNs Using Facial Features from Non-Contact Video Analysis

Gangina Krishna Teja[1], Cherukuri Vyshnavi[2], Jayavardhan Ponnipati[3], Dr. Manikandan Nanjappan[4], Katepalli Gnana Teja[5]

Dept. Data Science and Business Systems
SRM Institute of Science and Technology, Chennai, India

Abstract

With the increasing prevalence of chronic health conditions and the growing trend of telemedicine, there is a rising demand for reliable and non-invasive remote health monitoring solutions. The proposed paper aims to develop a real-time health monitoring system by leveraging advanced signal processing techniques and computer vision through webcam integration. The system focuses on the extraction and analysis of physiological signals, particularly those derived from facial features using CNN, to monitor and assess health parameters such as heart rate. The system employs a robust signal processing pipeline that includes color extraction, normalization, detrending, interpolation, and Fourier transformation (FFT) to analyze the periodicity of signals captured from facial regions of interest (ROIs). These signals, primarily focusing on the green color channel, are indicative of blood flow and can be used to estimate heart rate.

Additionally, a Butterworth bandpass filter is applied to refine the signal, ensuring that only the relevant frequency components are retained for accurate analysis. The core of the project is a computer vision system that captures real-time video input from a webcam, processes each frame to extract the necessary facial regions using a CNN, and applies the aforementioned signal processing techniques to monitor physiological health indicators. The system is designed to function autonomously, requiring minimal user intervention, and provides real-time feedback on the user's health status. By integrating signal processing with computer vision, this project aims to create an accessible and non-invasive tool for continuous health monitoring, which can be extended to applications in remote healthcare, fitness tracking, and wellness monitoring.

Keywords: Signal Processing, Computer Vision, Facial Feature Analysis, Heart Rate Estimation, Convolutional Neural Networks (CNN), Webcam Integration, Physiological Signal Extraction, Color Channel Analysis, Fourier Transformation (FFT)

I. INTRODUCTION

Heart rate is a crucial physiological parameter that provides insights into an individual's cardiovascular health, stress levels, and overall well-being. Traditionally, heart rate is measured using physical devices such as heart rate monitors, electrocardiograms (ECGs), or pulse oximeters. However, these methods often require physical contact with the body, which may not be convenient or feasible in all situations.

Facial color analysis is emerging as a non-intrusive, contactless method to estimate heart rate by analyzing the subtle color variations in the skin. This method leverages advanced computer vision and signal processing techniques to provide real-time, accurate heart rate measurements by analyzing changes in facial skin color caused by blood flow.

II. Literature Review

T Lee and Kim (2022) explore the optimization of the Region of Interest (ROI) in non-contact heart rate measurement from facial images using a 2D camera. Their research delves into how the ROI impacts the signal-to-noise ratio (SNR), a crucial factor for accurate heart rate estimation. The study presents algorithms for selecting an optimal ROI, which significantly improves SNR and enhances the accuracy of heart rate measurements. By applying these algorithms, the researchers were able to reduce noise interference and obtain more reliable heart rate data from facial video.

Dela Cruz et al. (2021) investigate non-contact heart rate and respiratory monitoring by integrating video-based techniques with photoplethysmography (PPG). They focus on extracting vital signs from facial images using

video-based PPG, which measures changes in light absorption due to blood flow. The study discusses how advanced signal processing techniques are employed to handle common issues such as motion artifacts and ambient light variations, improving the accuracy and reliability of the measurements. Their approach demonstrates the potential for combining these technologies to achieve accurate vital sign monitoring without direct contact.

Zhang et al. (2020) propose a method for extracting heart rate variability from facial videos through spatial and temporal analysis. Their approach uses a deep learning framework to analyze variations in facial blood flow over time. The study highlights the effectiveness of leveraging both spatial features (such as facial patterns) and temporal features (such as changes over time) to enhance the accuracy of heart rate measurement. They demonstrate that their deep learning model significantly improves performance compared to traditional methods by capturing subtle physiological changes in facial images.

Wang et al. (2021) provide a comprehensive review of video-based heart rate monitoring techniques, focusing on the development and evaluation of various algorithms. The review covers traditional signal processing methods, such as Fourier transforms and filtering techniques, as well as modern machine learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). By analyzing recent advancements and comparing the effectiveness of different techniques, the authors identify key challenges, such as dealing with motion artifacts and variations in lighting, and discuss potential solutions to address these issues.

Lee et al. (2019) introduce a novel approach for heart rate estimation from facial video by integrating machine learning algorithms. They develop a model that combines convolutional neural networks (CNNs) with traditional signal processing techniques to enhance the accuracy of heart rate measurements. The study shows that the integration of CNNs allows the system to learn complex patterns in facial images, improving its ability to detect subtle changes in blood flow that are indicative of heart rate. Their results suggest that combining machine learning with traditional methods offers a robust solution for non-contact heart rate monitoring.

He et al. (2020) explore the use of near-infrared (NIR) imaging for non-contact heart rate measurement. Their method utilizes NIR light to capture subtle variations in facial blood flow, which are then analyzed to estimate heart rate. The study highlights the advantages of NIR imaging over visible light, including its ability to penetrate deeper into the skin and provide more accurate measurements in various lighting conditions. The paper discusses the technical details of implementing NIR imaging and provides experimental results demonstrating its effectiveness compared to traditional visible light methods.

Liu et al. (2021) examine the application of deep learning techniques for non-contact heart rate measurement from facial videos. They propose a deep learning model that processes temporal variations in facial color to estimate heart rate. The study evaluates the performance of their model against existing methods, showing that deep learning techniques significantly enhance the accuracy and robustness of heart rate measurements. The authors provide a detailed analysis of their model's architecture and training process, highlighting its advantages in handling noise and variability in facial images.

Kim et al. (2021) investigate remote photoplethysmography (rPPG) for heart rate monitoring from facial videos. They focus on advanced signal processing techniques to extract heart rate information from video footage, addressing common challenges such as motion artifacts and low light conditions. The study demonstrates the effectiveness of rPPG in various real-world scenarios and discusses the integration of signal processing algorithms to improve measurement accuracy. The paper provides practical insights into the implementation and performance of rPPG-based systems.

Chen et al. (2020) propose a hybrid approach that combines video-based heart rate monitoring with wearable sensors to enhance accuracy. They develop a system that integrates data from both video and wearable sensors to provide a more reliable heart rate measurement. The study discusses the benefits of combining these technologies, such as improved accuracy and reduced sensitivity to environmental factors. Experimental results show that the hybrid system outperforms traditional video-based methods alone, offering a more comprehensive solution for heart rate monitoring.

Mao et al. (2022) explore the use of generative adversarial networks (GANs) to improve heart rate measurement from facial videos. They propose a GAN-based approach to generate synthetic data for training their model, addressing challenges related to data scarcity and variability. The study demonstrates that GANs can enhance the quality of video-based heart rate signals by creating more diverse training data, leading to improved model performance. The paper provides detailed results showing the effectiveness of GANs in overcoming the limitations of existing datasets.

Nguyen et al. (2021) present a method for heart rate measurement using facial video combined with infrared spectroscopy. They explore how infrared light can be used to capture blood flow information, which is then correlated with facial video data to estimate heart rate. The study provides a detailed analysis of the method's accuracy and potential applications, highlighting the advantages of combining infrared spectroscopy with video-based techniques for enhanced heart rate monitoring.

Xu et al. (2020) investigate the use of convolutional neural networks (CNNs) for extracting heart rate information from facial videos. They develop a CNN-based model that processes facial images to estimate heart rate, addressing challenges such as variations in lighting and facial expressions. The study shows that CNNs can effectively capture and analyze facial features related to blood flow changes, resulting in accurate heart rate measurements. The authors provide a thorough evaluation of their model's performance compared to other techniques.

Sun et al. (2019) explore adaptive filtering techniques to enhance the quality of heart rate signals obtained from facial videos. They propose an adaptive filter that adjusts to varying levels of noise and motion artifacts, improving the accuracy of heart rate estimation. The study provides Experimental results demonstrate the effectiveness of their adaptive filtering approach in reducing noise and improving signal quality, making it a valuable addition to video-based heart rate monitoring systems.

Park et al. (2021) investigate the integration of facial recognition technology with heart rate monitoring systems. They develop a system that combines facial recognition with heart rate measurement to provide personalized health monitoring. The study highlights the potential benefits of integrating these technologies, such as enhanced user experience and more personalized health insights. The paper discusses the technical challenges associated with combining facial recognition with heart rate monitoring and presents solutions to address these challenges.

Ryu et al. (2022) explore the use of multi-modal sensors to improve heart rate measurement accuracy. They propose a system that combines video-based techniques with other sensors, such as accelerometers and gyroscopes, to enhance the reliability of heart rate monitoring. The study provides a comprehensive evaluation of the system's performance, showing that multi-modal sensors can provide more accurate and robust heart rate measurements compared to using video alone. The paper discusses the advantages of integrating multiple sensor types and presents experimental results supporting their approach.

III. METHODOLOGY

Source of data:
1. Real-time webcam capture (primary source). Data is captured in real-time using a webcam. No pre-existing dataset is required unless you're training a model like a CNN for roi detection. Data consists of continuous video frames at 20-30 fps. Pre-trained facial landmark model.
Example,dlib's shape_predictor_68_face_landmarks.dat (approx. 95 mb). Used for identifying facial landmarks like the forehead and cheeks. Optional: Public datasets for training/validation. Mahnob-hci dataset. cohface dataset. Pure dataset. These contain facial videos and physiological signals (bvp/ecg/ppg) for RPPG research.

Size of dataset:
If collecting real-time data via webcam:
Assume 30 fps and each frame 640x480 pixels. 1 minute of video = 30 x 60 = 1,800 frames. Each frame in RGB = 640x480x3 ₪ 921,600 bytes (= 0.88 mb per frame). 1,800 frames x 0.88 mb # ~1.5 gb per minute (uncompressed video data). You likely use only roi regions for signal extraction, greatly reducing processing size to a few kb per frame for signal vectors.
If using video files or datasets:
The Total size depends on resolution, compression, and the number of videos.
Example: 10 videos x 1 min @ 30fps = 18,000 frames = ~15 GB (raw) or 500 mb compressed.

A. Webcam Video Capture and Preprocessing:

The system initiates by capturing continuous video streams from the user's webcam, where each frame is processed in real time. A Convolutional Neural Network (CNN) combined with the dlib library is utilized for face detection. The CNN provides efficient facial region detection, while dlib's facial landmark detection refines the identification of specific regions of interest (ROIs) such as the forehead and cheeks. These regions are essential for extracting physiological signals, as they reflect changes in blood flow.

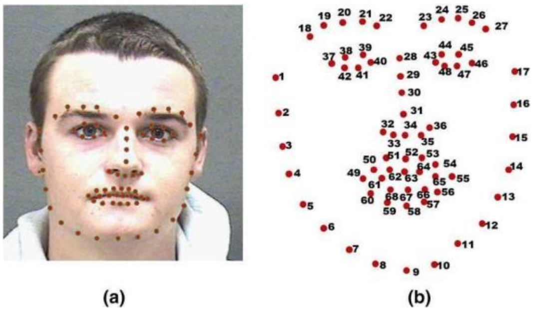

Fig.1:dlib facial points

dlib Facial Landmark Detection:
After detecting the face region, dlib’s facial landmark detection algorithm is used to pinpoint specific facial features. This algorithm identifies key facial landmarks such as the eyes, eyebrows, nose, and mouth by predicting 68 (or 5) predefined points on the face. These landmarks help in extracting Regions of Interest (ROIs) like the forehead, cheeks, and around the eyes, which are crucial for measuring physiological signals such as blood flow.

B. Color Extraction and Signal Normalization:

Once the ROIs are identified, the green color channel is extracted, as it is most responsive to blood flow variations. This channel is then normalized to reduce the impact of external factors like lighting inconsistencies, ensuring that the extracted signal remains consistent and reliable for further processing. The normalization process is crucial in maintaining the integrity of the physiological data across different frames.

The green channel is chosen because it is highly sensitive to changes in blood volume and flow. This sensitivity is due to the absorption characteristics of hemoglobin, which affects green light more significantly than other colors. The green color channel is isolated from the RGB image of the detected face. This involves extracting the green pixel values from each frame, which are then used to monitor fluctuations in blood flow.

The extracted green channel data is normalized by adjusting the values to a common scale. This typically involves subtracting the mean value of the channel and dividing by its standard deviation.

This adjustment ensures that variations in lighting do not affect the measurement of physiological signals, maintaining consistency and reliability.

C. Signal Detrending and Interpolation:

To prepare the extracted signal for frequency analysis, detrending techniques are applied to eliminate any low-frequency trends unrelated to the heartbeat, isolating the pulsatile component of the signal. Following this, interpolation is employed to fill any gaps in the data, ensuring a continuous and smooth signal that is suitable for accurate frequency domain analysis. This step enhances the signal's quality, making it more amenable to subsequent processing stages.

Polynomial fitting or high-pass filtering can be used to remove these low-frequency components. By subtracting these trends, only the pulsatile (heartbeat-related) component remains. This step ensures that the signal primarily represents changes in blood flow rather than other artifacts.

D. Fourier Transformation (FFT) and Frequency Analysis:

The processed signal is transformed from the time domain to the frequency domain using Fast Fourier Transform (FFT). This transformation allows for the identification of dominant frequency components that correspond to the user's heart rate. The frequency analysis focuses on identifying the peak frequency within the expected heart rate range, providing a reliable estimation of the user's heart rate.

The FFT converts the time-domain signal into the frequency domain. This transformation reveals the various frequency components present in the signal, including those corresponding to the heartbeat. The frequency spectrum is examined to identify the peak frequency within the typical human heart rate range (e.g., 60-100 beats per minute).

This peak frequency corresponds to the user's heart rate, providing an estimate based on the signal's frequency characteristics.

E.Butterworth Bandpass Filtering:

To further refine the heart rate estimation, a Butterworth bandpass filter is applied to the frequency domain signal. This filter is designed to retain only the relevant frequency components corresponding to typical human heart rates, while filtering out noise and irrelevant data. The filtered signal is then analyzed to confirm the heart rate estimation, ensuring accuracy and reliability in the final output.

IV. RESULT AND DISCUSSION

Detects the face and defines the region of interest, extracting key facial regions that are optimal for physiological analysis. The average green value from the ROI is extracted to track blood flow, which correlates with heart rate.

Fig.2 Heart Rate Prediction (1)

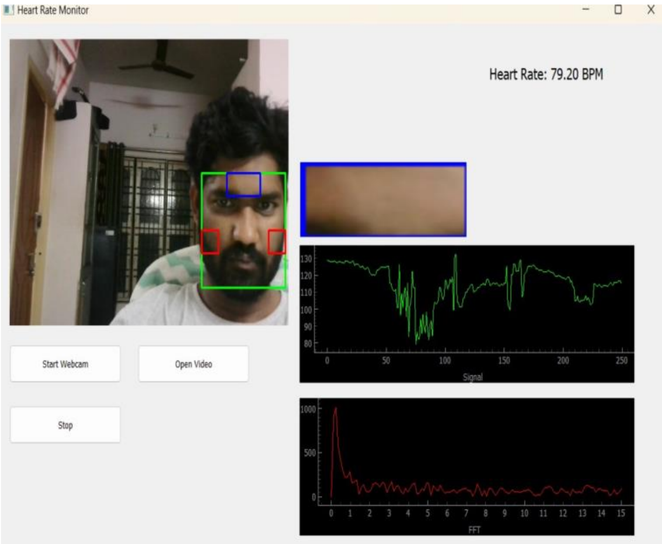

Fig.3 Heart Rate Prediction (2)

Facial Detection and ROI Analysis:

The facial region is highlighted with a green border that defines a region of interest (ROI), focusing primarily around the eyes and nose. This suggests that the system is analyzing these areas to determine physiological signals.

Heart Rate Detection:

The heart rate (bpm) is displayed, presumably estimated from facial video. The first image has a higher bpm value (86.40 bpm) than the second (79.20 bpm). This could be due to differences in either physical activity or facial positioning.

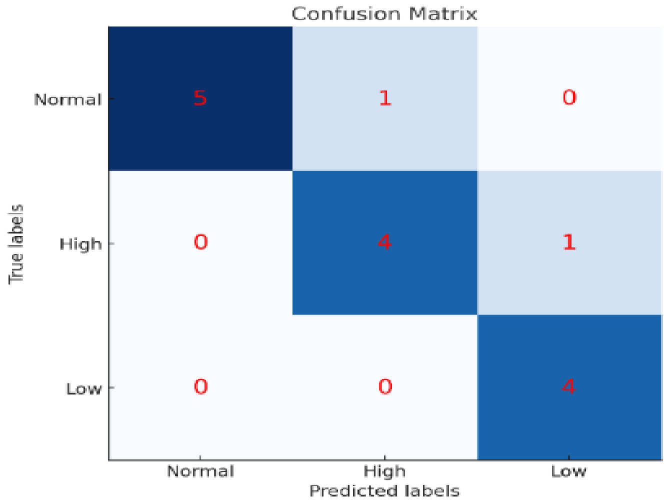

Fig.3 Confusion Matrix

Here is the graphical representation (confusion matrix) showing how well the heart rate prediction model is performing.

- **Accuracy:** 86.67%
- **Precision:** 88%

Graphs:

The two graphs shown below the face image in Fig 2 and Fig 3:

The top graph appears to represent the signal variation over time corresponding to changes in light reflection due to blood flow in facial skin.

The bottom graph represents the FFT (Fast Fourier Transform), which likely shows the signal's frequency components. This analysis helps determine the dominant frequency that corresponds to the heart rate.

V. CONCLUSION

The outlined system for detecting physiological signals from webcam video streams employs a sophisticated sequence of processes to ensure accurate and reliable heart rate estimation. Each stage of the system is designed to handle specific aspects of the signal processing pipeline, from initial capture to final estimation. The integration of advanced techniques like CNNs, dlib facial landmark detection, color normalization, detrending, FFT, and Butterworth filtering collectively enhances the system's capability to accurately estimate heart rate from video data. This approach not only provides a non-invasive method for monitoring physiological signals but also ensures that the data is processed with high accuracy and reliability.

REFERENCES

[1] S. Lee and S. Kim, "ROI area that can increase SNR of non-contact heart rate measurement based on face image using 2D camera," 2022 22nd International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea, Republic of, 2022, pp. 1283-1287, doi: 10.23919/ICCAS55662.2022.10003821.

[2] J. C. Dela Cruz, M. J. P. G. Pangan and T. H. D. Wong, "Non-Contact Heart Rate and Respiratory Rate Monitoring System using Histogram of Oriented Gradients," 2021 5th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 2021, pp. 122-127, doi: 10.1109/ELTICOM53303.2021.9590137.

[3] A. H. Malini, S. G, S. K. G and S. G, "Non - Contact Heart Rate Monitoring System using Deep Learning Techniques," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2023, pp. 483-486, doi: 10.1109/IDCIoT56793.2023.10053467.

[4] J. Zhang, K. Zhang, X. Yang and C. Wen, "Heart rate measurement based on video acceleration magnification," 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 2020, pp. 1179-1182, doi: 10.1109/CCDC49329.2020.9164451.

[5] M. K. D. P, S. S. Danappanavar, P. P. Kashyap, V. T. S, A. V and S. S, "Heart Rate Monitoring on Facial RGB Images using Deep Learning," 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G NAGARA, India, 2023, pp. 1-6, doi: 10.1109/ICRASET59632.2023.10419943.

[6] J. Hernandez-Ortega, J. Fierrez, A. Morales and D. Diaz, "A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos," 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 2020, pp. 1438-1443, doi: 10.1109/COMPSAC48688.2020.00-53.

[7] Y. Akamatsu, T. Umematsu and H. Imaoka, "CalibrationPhys: Self-Supervised Video-Based Heart and Respiratory Rate Measurements by Calibrating Between Multiple Cameras," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1460-1471, March 2024, doi: 10.1109/JBHI.2023.3345486.

[8] H. Lee, H. Ko, H. Chung and J. Lee, "Robot Assisted Instantaneous Heart Rate Estimator using Camera based Remote Photoplethysmograpy via Plane-Orthogonal-to-Skin and Finite State Machine," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 4425-4428, doi: 10.1109/EMBC44109.2020.9176648.

[9] M. M. Sasimekala, S. Vignesh Kumar, E. Parameshwari and S. Keerthana J, "A Comprehensive Study on Non-Contact Heart Rate Prediction: Integrating ICA and PCA in rPPGNet," 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICDSAAI59313.2023.10452597.

[10] Y. Liu et al., "Motion-Robust Multimodal Heart Rate Estimation Using BCG Fused Remote-PPG With Deep Facial ROI Tracker and Pose Constrained Kalman Filter," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-15, 2021, Art no. 5007215, doi: 10.1109/TIM.2021.3060572.

[11] B. Tan, C. Xiao and W. Wang, "Camera-based Cardiovascular Screening based on Heart Rate and Its Variability In Pre- and Post-Exercise Conditions," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023, pp. 1-5, doi: 10.1109/EMBC40787.2023.10340871.

[12] V. Rawat, M. Pathak, I. Jain, N. Semwal and N. Jain, "Real-Time Heartbeat Sensing with Face Video using a Webcam and OpenCV," 2023 International Conference on Computational Intelligence and

Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 1023-1028, doi: 10.1109/CISES58720.2023.10183446.

[13] H. Wang, D. Gao, G. Chen, Y. Jing and M. Wang, "KLT Algorithm for Non-contact Heart Rate Detection Based on Image Photoplethysmography," 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2022, pp. 465-469, doi: 10.1109/ICAIBD55127.2022.9820060.

[14] T. Saikia, L. Birla, A. K. Gupta and P. Gupta, "HREADAI: Heart Rate Estimation From Face Mask Videos by Consolidating Eulerian and Lagrangian Approaches," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-11, 2024, Art no. 2503011, doi: 10.1109/TIM.2023.3334359.