Current Journal of Engineering and Science Research
Volume 2, Issue 4, pp 1-6, April 2025
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2025.204001
Available Online at https://www.cjesr.com

VITASI: Real Time Remote Vital Health Monitoring

[1]Y Mohan Das, [2]K Meghana Reddy, [3]G Vamshi Krishna, [4]J Pavan Kumar, [5]G Yashwanth

[1,2,3,4,5]Department of Computer Science Engineering (Data Science), Gates Institute of Technology, Gooty, Andhra Pradesh, India
E-mail: [1]mohandassonu@gmail.com, [2]meghanareddy0421@gmail.com, [3]vamshikrishnayadav752@gmail.com, [4]pavanyadav30351@gmail.com, [5]yashwanthgajula7@gmail.com

Abstract: VITASI is a real-time remote health monitoring system that uses 3D imaging, multispectral, and thermal imaging to measure vital signs. It accurately tracks heart rate, breathing rate, oxygen saturation, and body temperature without physical contact. The system reduces errors by compensating for head movements and aligning different image types. It processes data at 15 frames per second, ensuring real-time monitoring. VITASI provides reliable results even with slight motion. The system can be expanded to measure additional health indicators like blood pressure. By integrating advanced imaging and AI, it offers a smart and efficient way to monitor health remotely. VITASI enhances contactless healthcare with high accuracy and real-time data processing.

Keywords: Vital signs, contactless monitoring, heart rate (HR), breathing rate (BR).

I. INTRODUCTION

Contactless vital sign measurement offers a hygienic and comfortable alternative to traditional contact-based methods, which can cause skin irritation, germ contamination, and movement restrictions. Image sensors enable non-invasive monitoring of heart rate, oxygen saturation, respiration rate, and body temperature. Heart rate estimation relies on photoplethysmographic (PPG) signals, analyzing skin color variations due to hemoglobin absorption. Oxygen saturation is determined using the "ratio of ratios" method, which compares pulsatile and nonpulsatile blood flow components at different wavelengths. Respiration rate is extracted from low-frequency PPG signals or thermal imaging, where temperature variations in the nostrils indicate breathing cycles. Additionally, thermal imaging helps estimate body temperature by detecting exhaled air warmth. However, challenges arise from motion artifacts, particularly head movement, which disrupts region-of-interest (ROI) tracking. Changes in ambient lighting further affect accuracy by introducing distortions in PPG signals. These factors necessitate advanced techniques.

To address these challenges, 3D imaging technology enables precise head motion tracking with six degrees of freedom, ensuring stable ROI selection. Active near-infrared (NIR) illumination minimizes lighting interference, as heart rate and oxygen saturation estimation are feasible in the 675–950 nm range. Multimodal imaging integrates 3D, NIR, thermal, and color imaging to enhance measurement reliability by combining multiple data sources.

A high-speed 3D sensor, aligned with 2D cameras, allows pixel-wise data registration for more accurate analysis. Novel algorithms process video data in real-time, significantly improving vital sign estimation. These advancements make it possible to overcome motion artifacts and lighting inconsistencies. The proposed system was successfully tested on subjects, demonstrating its effectiveness. By leveraging multimodal imaging and advanced signal processing, this approach offers a reliable contactless solution. It holds great potential for applications in medical.

II. RELATED WORK

Henault introduced a fully automatic, video-based framework for contactless heart rate monitoring, achieving high accuracy in near real-time using a webcam and laptop. Rizal developed a system-on-chip (SoC) device for measuring pulse and respiration rates, along with soft biometric parameters like age, gender, skin color, and height, showing high accuracy for some but limitations in age and gender estimation. Wang proposed a CNN-based framework combined with Phase-based Video Motion Processing (PVMP) to estimate heart and breathing rates, introducing a dataset with fewer constraints like movement and light variations. Rohmetra et al. explored machine learning-based vital sign monitoring using simple cameras and sensors, highlighting its potential for COVID-19 patients but also addressing challenges in implementation. Negishi et al.

Current Journal of Engineering and Science Research
Volume 2, Issue 4, pp 1-6, April 2025
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2025.204001
Available Online at https://www.cjesr.com

developed an RGB-thermal sensor fusion approach for rapid heart rate, respiration rate, and body temperature measurements, demonstrating its effectiveness in infection screening.and offline, based on a CNN for HR and BR estimations.

Tran et al. built a deep-learning-based system using object detection (YOLOv3) and signal processing to estimate breathing rate, heart rate, and blood pressure, meeting medical standards for accuracy. Yang et al. designed a contactless monitoring system using thermal and RGB cameras to measure body temperature, heart rate, and respiration rate, proving effective even for individuals wearing face masks. Selvaraju et al. reviewed data acquisition technologies, emphasizing the need for robust algorithms and fusion-based approaches for reliable contactless vital sign monitoring. With the growing interest in digital health, smartphone camera-based vital sign measurement could be a promising solution for remote healthcare applications.

III. PROPOSED SYSTEM

In proposed work, WEBCAM will read 10 frames which contains human face and then detect face and then extract ROI (region of interest) of face and then extract temporal (current time data) and spatial (current frame data) features and then input this features to Phase-based Video Motion Processing (PVMP) algorithm which will extract PPG signals and then these signals will be input to CNN model to predict heart and breath rate.

In this project we have used COHFACE dataset which contains 11 person videos and from this videos we extracted faces, heart and breathe rate and input to CNN algorithm to train model.

All this features will be calculated from WEBCAM and to implement this project we have designed following modules:

1) Generate & Load VitaSi CNN Model: using this module we will load CNN model which will predict heart and breath rate.
2) Contactless Vital Estimation: this module will open WEBCAM and then read 10 frames and then estimate heart and breathe rate by extracting PPG signal and employing CNN model.
3) MSE Graph: using this module we will display MSE (mean square error) of CNN model for heart and breathe rate prediction. The lower the MSE the better is the prediction model.

PPG Signals

PPG (Photoplethysmography) signals are a type of physiological signal that measures changes in blood volume in peripheral blood vessels. PPG is a non-invasive method used to monitor various cardiovascular parameters, such as heart rate, blood oxygen saturation, and pulse waveforms. The PPG signal is obtained by illuminating the skin with a light source, typically an LED, and detecting the light that is transmitted or reflected back using a photodetector. The light is absorbed differently by oxygenated and deoxygenated blood, leading to variations in the intensity of the detected light. These intensity changes in the PPG signal represent the pulsatile component of blood volumechanges associated with each heartbeat.

Phase based video motion processing

Phase-based video motion processing is a technique used for analyzing and understanding motion in video sequences. It is based on the idea that motion information can be extracted from the phase component of the video signal. This approach has gained popularity due to its ability to handle various types of motion, including translational, rotational, and complex motions. The process starts by decomposing a video sequence into its constituent frames. Each frame is then transformed into the frequency domain using techniques such as the Fourier transform or the Gabor transform. The phase component of the transformed frames contains information about the local phase shifts in the video signal.

CNN Model

The training and testing of a Convolutional Neural Network (CNN) in VITASI involve processing facial motion data through multiple convolution layers, followed by activation functions like the Rectified Linear Unit (ReLU), max pooling layers, fully connected layers, and a SoftMax classification layer. The SoftMax layer assigns probabilistic values to classify the extracted features effectively.

Current Journal of Engineering and Science Research
Volume 2, Issue 4, pp 1-6, April 2025
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2025.204001
Available Online at https://www.cjesr.com

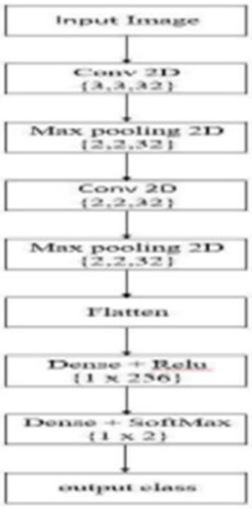


Figure 1: Representation of convolution layer process

The convolution layer is the fundamental component responsible for feature extraction. It preserves spatial relationships by applying a filter (kernel) to small regions of the input data. Mathematically, the convolution operation processes an input image,wherespatial coordinates represent rows and columns, and the depth represents the number of channels. A kernel of fixed dimensions slides over the input data to compute feature maps. The resulting feature map highlights essential patterns relevant to vital sign estimation.

The ReLU activation function is applied to introduce non-linearity into the network, allowing it to capture complex patterns in facial motion. This function outputs the input value if it is positive; otherwise, it returns zero, thereby enhancing model efficiency and convergence speed.

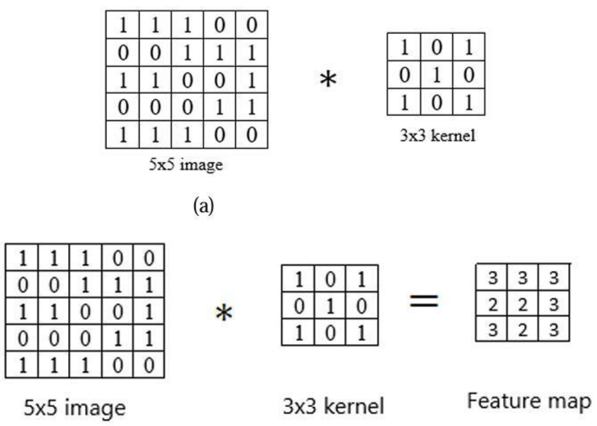


Fig. 2: Example of convolution layer process. (a) a speech with size kernel. (b) Convolved feature map

Max pooling is used to reduce the number of parameters and computations in the network by downsampling feature maps while retaining the most significant information. It helps in reducing dimensionality while ensuring that critical spatial features remain intact for further processing.

SoftMax classifier

The SoftMax function is utilized at the final stage of the convolutional neural network (CNN) to classify and interpret extractedfeatres from facial motion data. As the last layer of the model, SoftMax ensures that all the nodes converge, and the most probable classification is selected based on the processed input.

The input to the model is derived from real-time facial motion captured by a webcam. The hidden layers process this data, extracting vital patterns and features related to physiological parameters. The final processed data is then passed to the SoftMax layer, which determines the classification probabilities. If the model is trained to recognize different physiological states, each class represents a specific health condition or vital sign category.

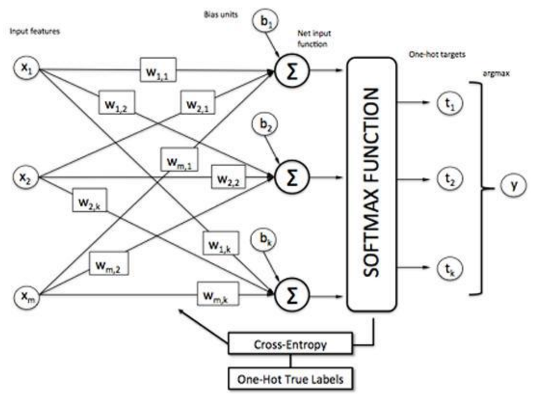


Fig. 3: SoftMax classifier.

For example, if the system is designed to classify heart rate variations into different ranges, each range is assigned a specific output node. SoftMax assigns a probability to each node, and the one with the highest probability determines the final classification. The function relies on a technique called One-Hot Encoding, which assigns a binary vector to each class. If SoftMax predicts a heart rate category, it assigns a probability distribution across all classes, ensuring the highest probability corresponds to the correct classification.

Current Journal of Engineering and Science Research
Volume 2, Issue 4, pp 1-6, April 2025
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2025.204001
Available Online at https://www.cjesr.com

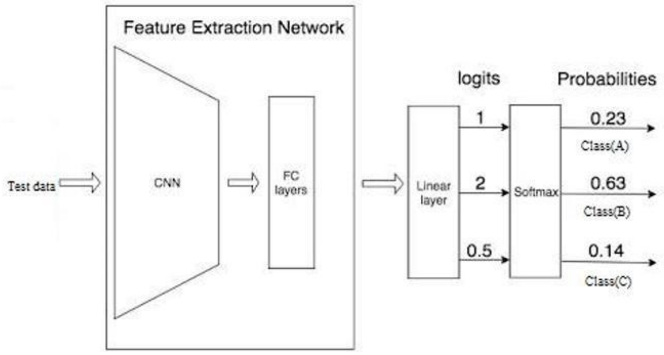


Fig. 4: Example of SoftMax classifier.

To improve accuracy, the system minimizes the classification error using a cross-entropy lossfunction. This function calculates the difference between the predicted and actual class, guiding the model toward better predictions. The lower the loss value, the more accurate the model's classification. Mathematically, the loss function is computed as:LOSS = np.sum(-Y * np.log(Y_pred))

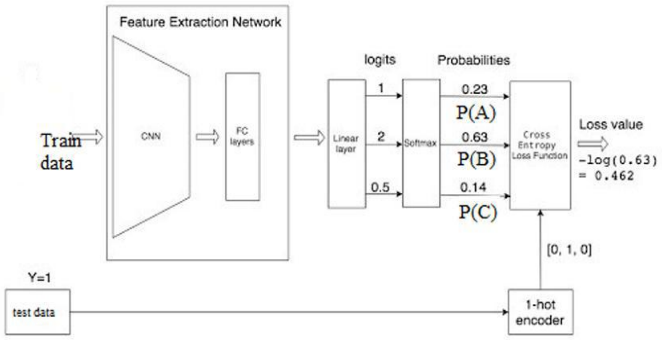


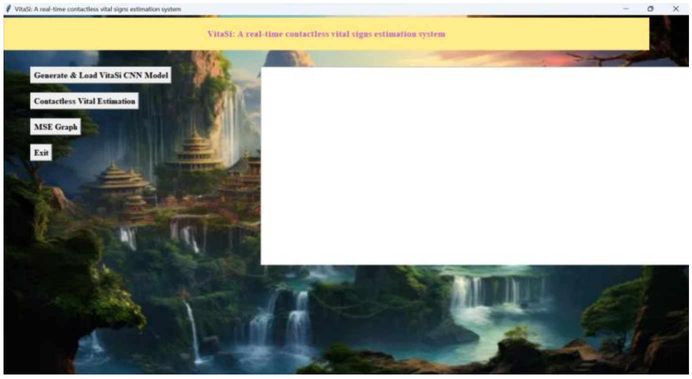
Fig. 5: Example of Soft Max classifier with test data

By integrating SoftMax classification, the VITASI system can accurately estimate heart rate and respiration rate, ensuring real-time remote health monitoring without direct physical contact. This makes it a valuable tool in healthcare applications, enabling efficient and accurate physiological assessments.

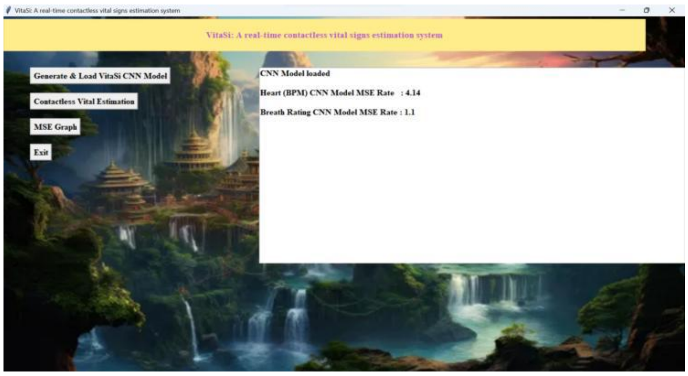
IV. RESULTS

In this project we have used COHFACE dataset which contains 11 person videos and from these videos we extracted faces, heart and breathe rateand input to CNN algorithm to train model. All these features will be calculated from WEBCAM and to implement this project we have designed following modules. Generate & Load VitaSi CNN Model: using this module we will load CNN model which will predict heart and breath rate. Contactless Vital Estimation: this module will open WEBCAM and then read 10 frames and then estimate heart and breathe rate by extracting PPG signal and employing CNN model. MSE Graph: using this module we will display MSE (mean square error) of CNN model for heart and breathe rate prediction. The lower the MSE the better is the prediction model.

The below screen click on ‘Generate & Load VitaSi CNN Model’ button to generate and load CNN model and get below screen.

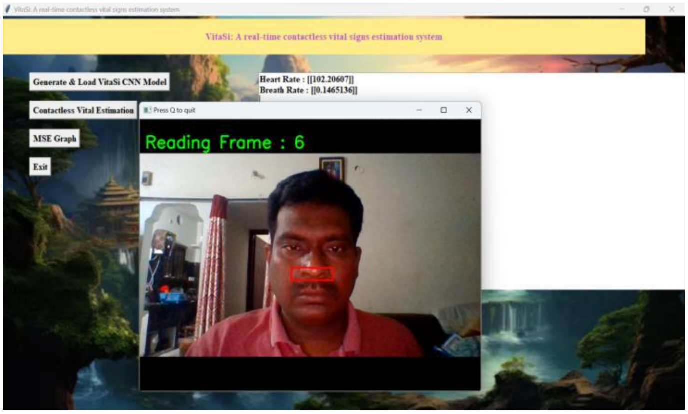


In above screen WEBCAM read 10 frames and then predict heart and breathe rate and this rate will be displayed and updated on TEXT AREA and in above screen we got heart rate as 106 and breathe rate as 0.13 and now click on ‘MSE Graph’ button to get below graph.

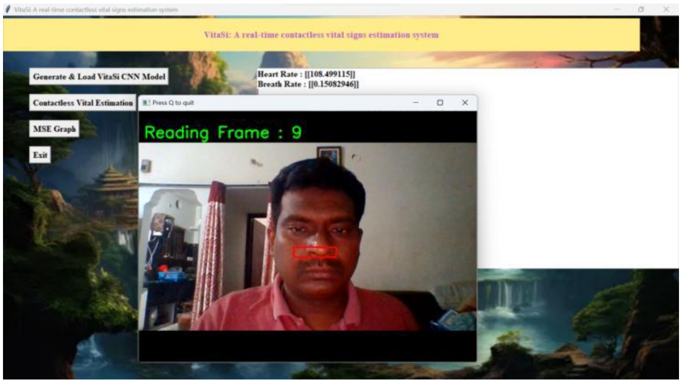


In above screen CNN model is loaded and we got Heart Rate MSE as 8.11 and Breathe Rate MSE as 2.62 and now click on ‘Contactless Vital Estimation’ button to start WEBCAM and predict Heart Rate and Breathe Rate.

Current Journal of Engineering and Science Research
Volume 2, Issue 4, pp 1-6, April 2025
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2025.204001
Available Online at https://www.cjesr.com



In above screen WEBCAM read 10 frames and then predict heart and breathe rate and this rate will be displayed and updated on TEXT AREA and in above screen we got heart rate as 106 and breathe rate as 0.13 and now click on 'MSE Graph' button to get below graph.



In below screen WEBCAM read 10 frames and then predict heart and breathe rate and this rate will be displayed and updated on TEXT AREA and in above screen we got heart rate as 106 and breathe rate as 0.13 and now click on 'MSE Graph' button to get below graph.

In above graph x-axis contains type of prediction and y-axis contains MSE error value of CNN prediction.

V. CONCLUSION

The estimated oxygen saturation values show some noise due to uneven illumination and spectral crosstalk in NIR cameras. To improve accuracy, we plan to enhance the LED array for uniform lighting and use narrower band-pass optical filters. Additionally, investigating correlations between vital signs, such as heart rate and respiration, may help refine estimations. This study serves as a demonstration, with future clinical evaluations planned to compare results with accurate contact-based measurements.

REFERENCES

[1] Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote plethysmographic imaging using ambient light. Opt. Express 2008, 16, 21434–21445.
[2] De Haan, G.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. IEEE Trans. Biomed. Eng. 2013, 60, 2878–2886.
[3] Xu, S.; Sun, L.; Rohde, G.K. Robust efficient estimation of heart rate pulse from video. Biomed. Opt. Express 2014, 5, 1124–1135.
[4] Rapczynski, M.; Werner, P.; Saxen, F.; Al-Hamadi, A. How the Region of Interest Impacts Contact Free Heart Rate Estimation Algorithms. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2027–2031.
[5] Tarassenko, L.; Villarroel, M.; Guazzi, A.; Jorge, J.; Clifton, D.A.; Pugh, C. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. Physiol. Meas. 2014, 35, 831–907.
[6] Kumar, M.; Veeraraghavan, A.; Sabharwal, A. DistancePPG: Robust non-contact vital signs monitoring using a camera. Biomed. Opt. Express 2015, 6, 1565–1588.
[7] Guazzi, A.R.; Villarroel, M.; Jorge, J.; Daly, J.; Frise, M.C.; Robbins, P.A.; Tarassenko, L. Non-contact measurement of oxygen saturation with an RGB camera. Biomed. Opt. Express 2015, 6, 3320–3338.
[8] Scully, C.G.; Lee, J.; Meyer, J.; Gorbach, A.M.; Granquist-Fraser, D.; Mendelson, Y.; Chon, K.H. Physiological Parameter Monitoring from Optical Recordings with a Mobile Phone. IEEE Trans. Biomed.

Current Journal of Engineering and Science Research
Volume 2, Issue 4, pp 1-6, April 2025
ISSN (online): 3049-1738
DOI: https://doi.org/10.47001/CJESR/2025.204001
Available Online at https://www.cjesr.com

Eng. 2012, 59, 303–306.
[9] Bal, U. Non-contact estimation of heart rate and oxygen saturation using ambient light. Biomed. Opt. Express 2015, 6, 86–97.
[10] Rosa, A.; Betini, R. Noncontact SpO2 Measurement Using Eulerian Video Magnification. IEEE Trans. Instrum. Meas. 2020, 69, 2120–2130.
[11] Van Gastel, M.; Stuijk, S.; De Haan, G. Robust respiration detection from remote photoplethysmography. Biomed. Opt. Express 2016, 7, 4941–4957.

Citation of this Article:

Y Mohan Das, K Meghana Reddy, G Vamshi Krishna, J Pavan Kumar, & G Yashwanth. (2025). VITASI: Real Time Remote Vital Health Monitoring. *Current Journal of Engineering and Science Research.* 2(4), 1-6. Article DOI: https://doi.org/10.47001/CJESR/2025.204001

***** End of the Article *****