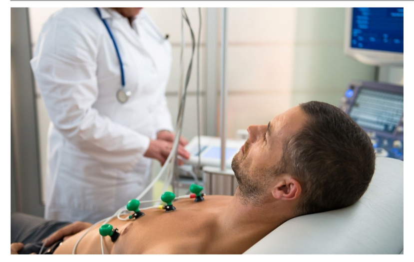
Developing a Non-Contact Cardiac Monitoring System via Remote Photoplethysmography (rPPG)

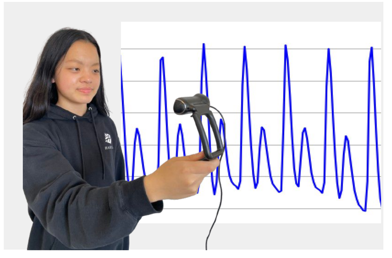
Ellie Zhou

Issue Statement

1. Cardiovascular diseases are the **leading cause of death** worldwide, claiming over **18 million** lives annually.
2. Rural populations face nearly **twice** the mortality rate of urban populations due to limited healthcare access.

Electrocardiogram (ECG):	**Photoplethysmography (PPG) from pulse oximeter**	**Remote Photoplethysmography (rPPG):**
- Clinically proven - Needs electrodes attached to skin - Requires visit to medical facility - Inaccessible for elderly, remote communities	- Requires direct skin contact - Uncomfortable - Sensitive to skin tone, motion artifacts	- Contactless - Low-cost - Easy to collect - Noisy and inaccurate





Background

Photoplethysmography (PPG) - measures changes in blood volume

- LED emits light onto skin. Blood absorbs some light, rest is reflected back. Photodetector captures reflected light, which fluctuates with heart beat
- Systole (heart beat) ⟶ Blood volume increase ⟶ More light absorbed, less light reflected
- Diastole (between beat) ⟶ Blood volume decrease ⟶ Less light absorbed, more light reflected

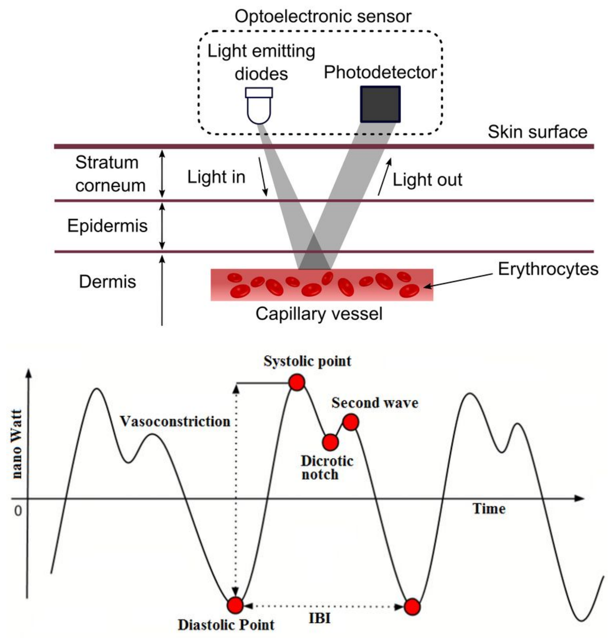


Image Credits: Research Gate

Remote Photoplethysmography (rPPG)

- Extracts rPPG signal from light reflections captured by a simple webcam
- Susceptible to noises from the video, limiting its real-world applicability

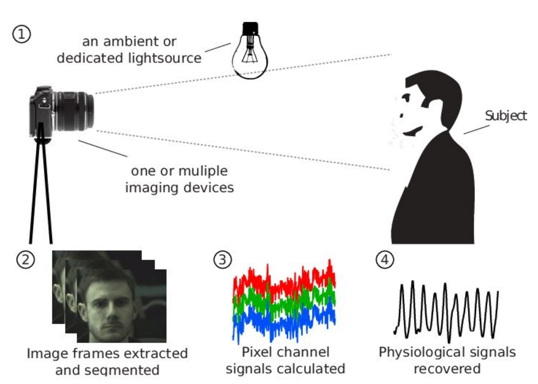


Image Credits: Research Gate

Research Question and Objective

Question	How can I design a deep learning solution to acquire an accurate rPPG signal and obtain effective cardiovascular metrics?
Objective	● Develop a solution to effectively acquire the rPPG signal ● Address challenges in rPPG accuracy and reliability

Related Work

2D CNN	● Models: DeepPhys [2], TS-CAN [3], EfficientPhys [4] ● Extracts Blood Flow Signal from Spatial Features ● **Gap:** does not learn temporal changes
LSTM	● Models: PhysNet-LSTM [7] ● Effectively models temporal sequences ● **Gap:** struggles with abrupt noise, computationally expensive
3D CNN	● Models: PhysNet-3DCNN [7] ● Simultaneous spatial and temporal feature learning ● **Gap:** PPG signal fails in rotation settings

Overall System Architecture

Acquire Facial Video ➡ Preprocess Video frames ➡ Skin Segmentation ➡ Model with Intelligence Branch ➡ rPPG Output ➡ Heart Metrics

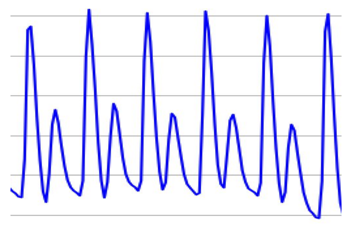
Acquire Facial Video
2 datasets:
- UBFC-rPPG
- MMPD

Preprocess Video frames
- Crop Face
- Resize frames to 72x72
- Chunk to 128 frames

Skin Segmentation
- Extract skin region
- Remove irrelevant features that introduce noise

Model with Intelligence Branch
- 3D-Convolutional Neural Network
- Create attention masks to promote stable rPPG signal

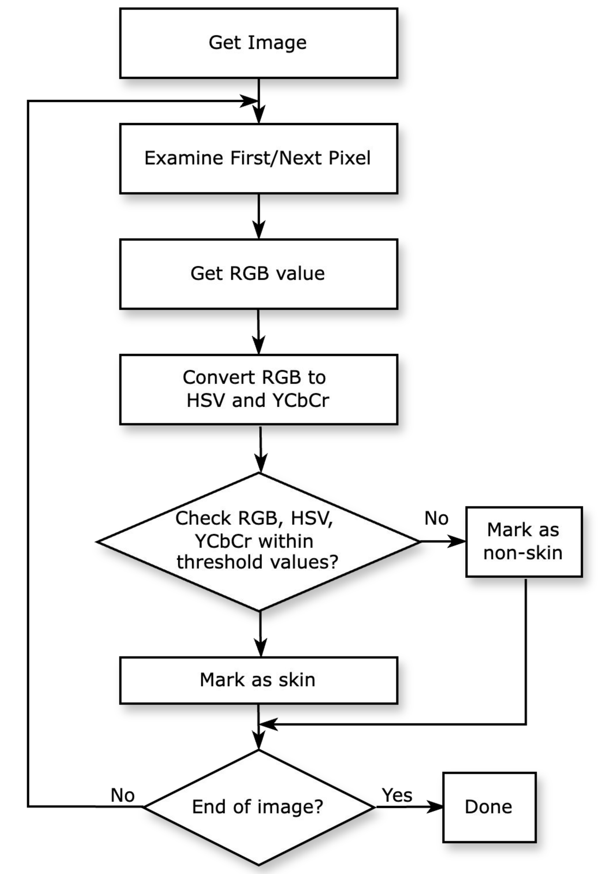
rPPG Output



Heart Metrics

Architecture Innovations

1. Skin Segmentation



Color Space	Thresholds
RGB	R > 95 and G > 40 and B > 20 and R > G and R > B and | R - G | > 15
HSV	0.0 <= H <= 50.0 and 23% <= S <= 68%
YCbCr	Cb > 85 and Y > 80 Cr <= (1.5862*Cb)+20 and Cr >=(0.3448*Cb)+76.2069 and Cr >= (-4.5652*Cb)+234.5652 and Cr <= (-1.15*Cb)+301.75 and Cr <= (-2.2857*Cb)+432.85

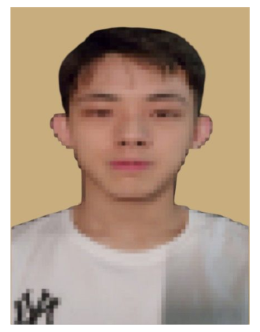


Fig. 1: Image from video


Fig. 2: Skin Mask

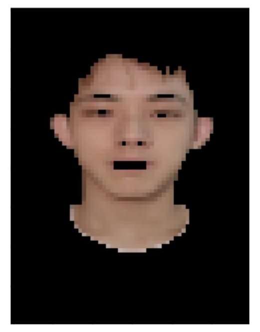
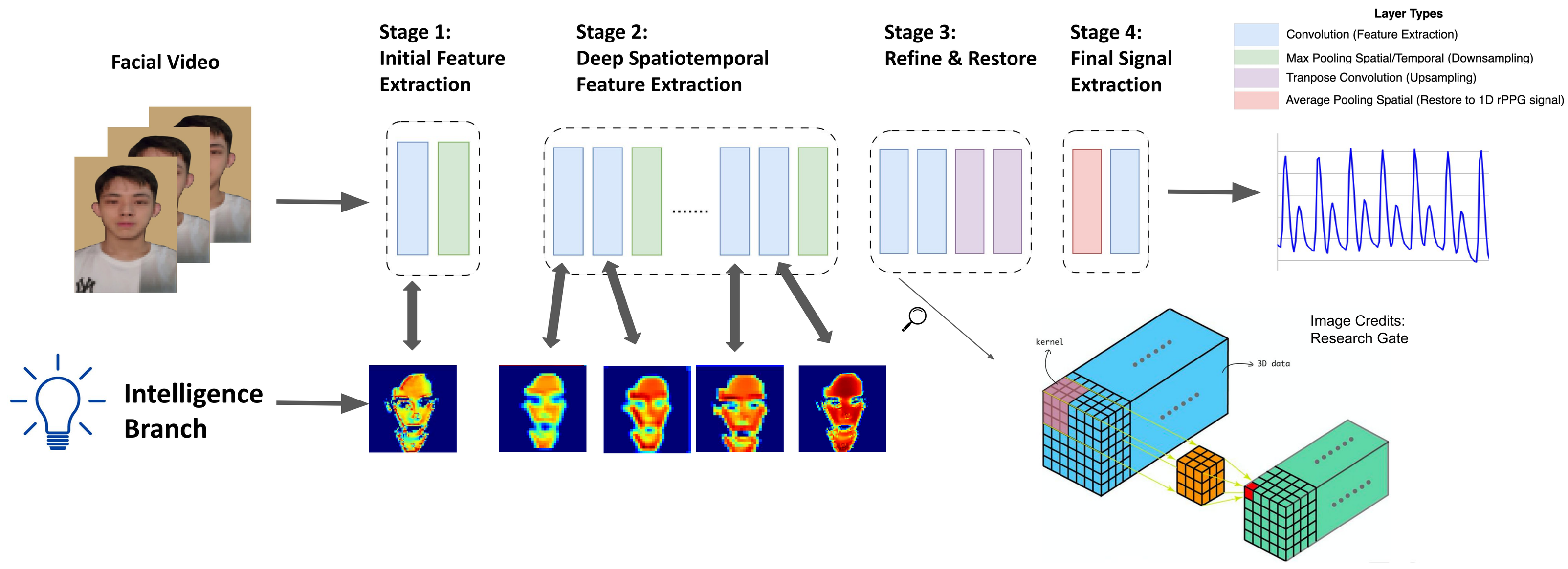


Fig 3: Segmented Image

2. 3D CNN



3. Intelligence Branch

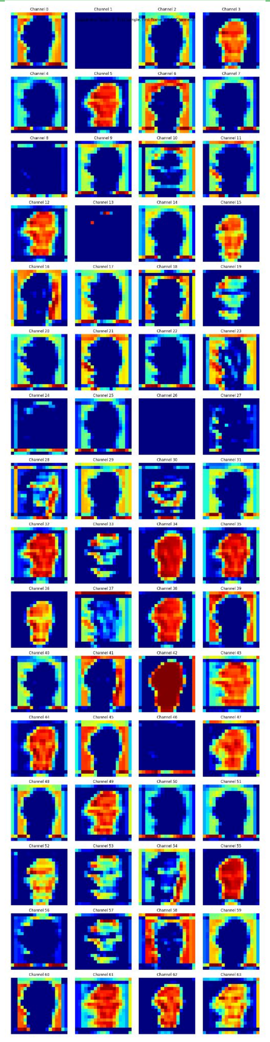
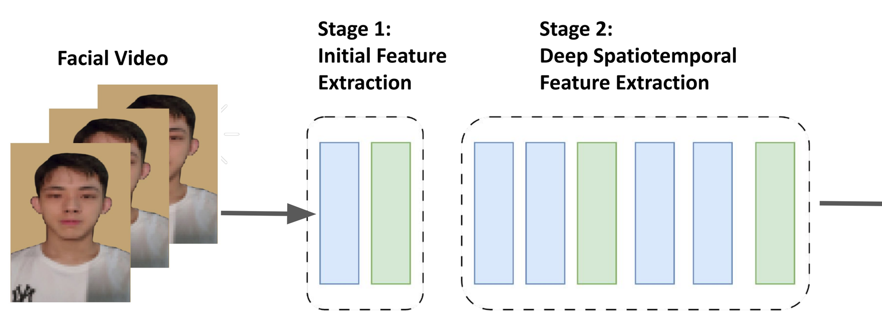


Fig. 6: 64 channels



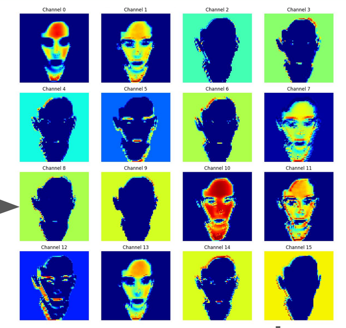
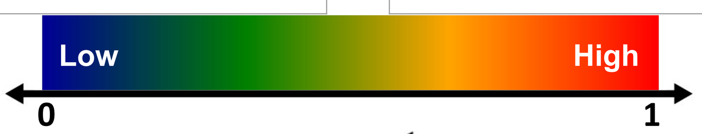


Fig. 4: 16 channels

Attention Mask Values (High vs. Low)

Low	High
● High motion areas (talking mouth, side of the face during head turn) ● Shadowed regions (side of nose, near hairline)	● Regions with minimal motion (forehead) ● Skin regions where blood volume changes are more visible (cheeks)



$$A = \frac{1}{1+e^{-x}}$$

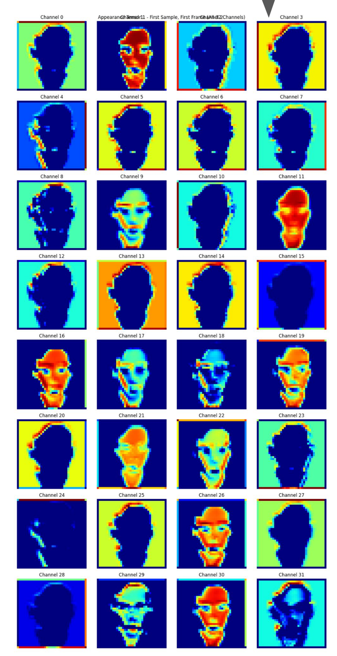


Fig. 5: 32 channels

Results

Table 1 - Dataset: UBFC-rPPG (controlled conditions)					
MAE	**PhysNet**	**DeepPhys**	**EfficientPhys**	**TS-CAN**	**My Model**
Heart Rate	1.465	5.713	2.490	2.490	0.586

Table 2 - Dataset: MMPD Stationary with various lighting, skin tone					
MAE	**PhysNet**	**DeepPhys**	**EfficientPhys**	**TS-CAN**	**My Model**
Heart Rate	3.343	23.599	20.551	7.559	1.993

Table 3 - Dataset: MMPD Rotation with various lighting, skin tone					
MAE	**PhysNet**	**DeepPhys**	**EfficientPhys**	**TS-CAN**	**My Model**
Heart Rate	5.399	24.346	21.772	28.257	4.049

Table 4 - Dataset: MMPD Talking with various lighting, skin tone					
MAE	**PhysNet**	**DeepPhys**	**EfficientPhys**	**TS-CAN**	**My Model**
Heart Rate	3.164	26.235	18.062	3.999	3.547

Table 5 - Other Heart Metrics					
MAE	**PhysNet**	**DeepPhys**	**EfficientPhys**	**TS-CAN**	**My Model**
IBI	8.201	176.786	117.551	134.719	7.355
SDNN	16.423	80.522	51.223	73.196	14.982
SDSD	18.638	67.798	49.569	63.813	15.624
RMSSD	28.584	124.099	82.693	112.529	20.984
SD1	20.189	87.521	58.334	79.531	14.837
SD2	15.475	57.853	44.225	68.951	11.483
Breathing Rate	0.014	0.022	0.011	0.020	0.023

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x - x_i|$$

x = true value
x_i = prediction value
n = # samples

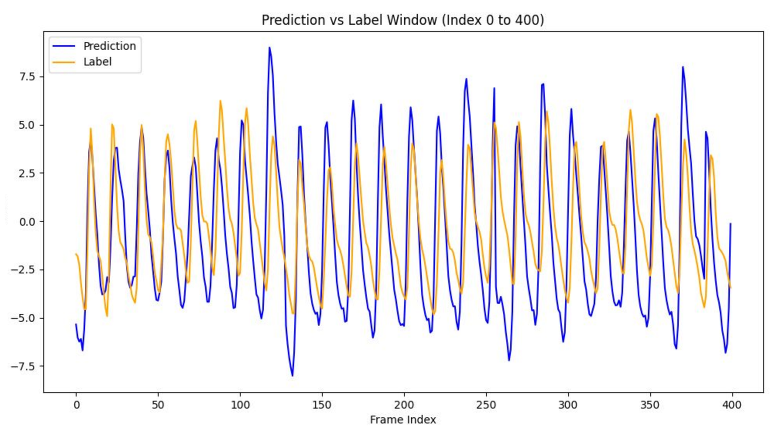


Fig. 7

Analysis and Discussion

1. Table 1: In controlled settings, my model attains MAE of 0.586 beats per minute (bpm), outperforming other models by 0.88 to 5.13 bpm. My model predicts heart rate with an error of less than 1 bpm, approaching medical-grade accuracy in a non-contact method.
2. Table 2: In the MMPD Stationary dataset, which introduces variations in skin tone and lighting without motion artifacts, my model achieves the lowest MAE of 1.993 bpm, outperforming state-of-the-art models by 1.35 to 21.61 bpm. This shows my model's superior ability to generalize across diverse skin tones and illumination conditions.
3. Table 3: In the MMPD Rotation dataset, my model achieves MAE of 4.049 bpm. My model solves the problem of rotation, significantly outperforming existing state-of-the-art models.
4. Table 4: In the MMPD Talking dataset, my model achieves MAE of 3.547 bpm, performing close to PhysNet. My model is robust to facial motion artifacts.
5. Table 5: My model is capable of accurately measuring other key cardiac metrics such as IBI, SDNN, and SDSD.
6. Figures 4-6: In the Intelligence Branch outputs, the high attention values at forehead and cheeks indicate that the model prioritizes high-perfusion regions, where the rPPG signal is strongest.
7. Figure 7: The predicted PPG follows the same shape as the ground truth. The signal is smooth, demonstrating my model's strong ability to denoise and mitigate motion artifacts.
8. Limitations: limited access to high-performance computing, access to university/research institution datasets, and time for model training

Real-World Applications

- Opens up new avenues in telemedicine by enabling remote diagnostics and timely intervention for cardiovascular conditions.
- Enables continuous home-based monitoring for the elderly. Improves management of heart health.
- Provides vital heart health insights to people in remote and rural areas.
- Facilitates contactless health monitoring for neonatal babies, ensuring safe and unobtrusive observation of vital signs.

Future Work

- Improve rPPG signal extraction on talking motion scenarios.
- Apply dimensionality reduction to improve computational efficiency.
- Develop multimodal models integrating rPPG with additional physiological signals and health data for diagnosis of cardiovascular diseases such as Atrial Fibrillation, Arrhythmias, and Heart Failure.

Impact

Skin Segmentation
Accurately extracts skin regions, removing noise to ensure that only relevant facial areas are used for rPPG.

3D CNN
Learn both spatial and temporal features. Effectively mitigates motion artifacts and variations in lighting and skin tone.

Intelligence Branch
Dynamically guides the model to focus on the highest quality rPPG signal regions.

Cardiac Metrics
Empowers the patients to gain insight into their cardiovascular health and potential risks.

References

1. Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., Dubois, J. (2017) "Unsupervised skin tissue segmentation for remote photoplethysmography", Pattern Recognition Letters
2. Chen, W., & McDuff, D. (2018). *DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks*. Proceedings of the European Conference on Computer Vision (ECCV), 2018. Retrieved from https://openaccess.thecvf.com/content_ECCV_2018/papers/Weixuan_Chen_DeepPhys_Video-Based_Physiological_ECCV_2018_paper.pdf
3. Liu, X., Fromm, J., Patel, S., & McDuff, D. (2020). *Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement*. arXiv preprint arXiv:2006.03790. Retrieved from https://arxiv.org/abs/2006.03790
4. Liu, X., Hill, B., Jiang, Z., Patel, S., McDuff, D. (2023). *EfficientPhys: Enabling Simple, Fast, and Accurate Camera-Based Cardiac Measurement*. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023. Retrieved from https://openaccess.thecvf.com/content/WACV2023/papers/Liu_EfficientPhys_Enabling_Simple_Fast_and_Accurate_Camera-Based_Cardiac_Measurement_WACV_2023_paper.pdf
5. Solina, F., Peer, P., Batagelj, B., Juvan, S., & Kovac, J. (2003). *Color-Based Face Detection in the "15 Seconds of Fame" Art Installation*. Retrieved from http://lrv.fri.uni-lj.si/~peterp/publications/mirage03.pdf
6. Tang, J., Chen, K., Wang, Y., Shi, Y., Patel, S., McDuff, D., Liu, X. (2023). *MMPD: Multi-Domain Mobile Video Physiology Dataset*. Proceedings of the IEEE in Engineering in Medicine and Biology Society. Retrieved from https://arxiv.org/pdf/2302.03840
7. Yu, Z., Li, X., Zhao, Y., & Zhao, G. (2019). *Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks*. arXiv preprint arXiv:1905.02419. Retrieved from https://arxiv.org/abs/1905.02419