



Improving biomedical diagnosis
through light-based technologies
and machine learning

Doctoral Networks (MSCA-DN)

HORIZON - MSCA - 2022 - DN

Grant Agreement N° 101119924

WP2: Cardiovascular diseases

Deliverable D2.1
Preprocessing algorithms

Muhammad Usman, PG Gdansk

Date: 30.09.2025

Type:

OTHER/R — Document, report.

Dissemination Level: PU-Public



Funded by
the European Union

DELIVERABLE IDENTIFICATION SHEET

Document version	Date	Editor	Change(s)
1	04-08-2025	Muhammad Usman	First version
2	20-08-2025	Jacek Rumiński	Second version
3	22-08-2025	Muhammad Usman	Third version

Grant Agreement N°	101119924
Call Identifier	HORIZON - MSCA - 2022 – DN
Acronym	BE-LIGHT
Full title	Improving biomedical diagnosis through light-based technologies and machine learning
Project URL	http://www.belightproject.eu/

Deliverable N° and Name	D2.1 Preprocessing algorithms
Work Package	WP2 <i>Cardiovascular diseases</i>
Lead Beneficiary	PG
Type	OTHER/R – Document, report.
Dissemination Level	PU-Public
Due date	Month 24

Authors (Partner)	Muhammad Usman, PG; Jacek Rumiński, PG		
Responsible Author	Muhammad Usman	Email	muhammad.usman1@pg.edu.pl
	Partner PG	Phone	

Abstract (for dissemination)	Cardiovascular disease (CVD) is one of the leading causes of mortality worldwide, emphasizing the need for continuous, scalable and non-invasive monitoring. Although traditional methods, such as ECG and PPG, are effective but have limitations because they require direct skin contact and are used in clinical settings. To address these challenges, we propose STREAM-Net, a dual-branch deep learning framework for remote photoplethysmography (rPPG), enabling contactless vital sign measurement through facial video analysis. Our approach incorporates preprocessing techniques and Monte Carlo dropout for physiological signal enhancement and uncertainty quantification to ensure robust and reliable predictions. Furthermore, we explore advanced image enhancement methods, including super-resolution and deblurring (e.g., ESDR, LapSRN, and FaceSR) to enhance signal recovery in both RGB and thermal imagery. Experimental results demonstrate that enhanced image quality significantly improves rPPG feature extraction.
Keywords	Cardiovascular disease, remote photoplethysmography (rPPG), super resolution, remote health monitoring, machine learning

Table of Contents

DELIVERABLE IDENTIFICATION SHEET	2
EXECUTIVE SUMMARY.....	4
INTRODUCTION	4
1 PREPROCESSING FOR STREAM-NET	6
1.1 Key Preprocessing Steps	6
2 MC-DROPOUT FOR UNCERTAINTY ESTIMATION.....	7
2.1 Implementation.....	8
2.2 MC-Dropout Predictions and Visualization	8
3 SUPER-RESOLUTION (SR) ALGORITHMS	13
3.1 Experimental details	13
3.2 Methods Evaluated	14
3.3 Results and Evaluation	15
4 FUTURE WORK FOR D2.6.....	17
REFERENCES	17
ANNEX	19

EXECUTIVE SUMMARY

Cardiovascular disease (CVD) is the leading global cause of mortality, emphasizing the need for continuous, scalable, and non-invasive monitoring. Although Traditional methods, such as ECG and PPG, are effective, but limited by their reliance on direct contact with skin and clinical settings. To address these limitations, we proposed STREAM-Net [14], a dual-branch deep learning framework for remote photoplethysmography (rPPG), which measures vital signs simply by analysing facial videos. This approach eliminates the need for physical contact or specialized equipment. Our proposed STREAM-Net employs a preprocessing pipeline, which incorporates resizing, face cropping, and temporal normalization to ensure robust spatio-temporal feature extraction. Additionally, we integrated a Monte Carlo dropout technique to quantify uncertainty in predictions to measure reliability and its applicability in clinical simulation. We further extend our research to investigate how advanced image enhancement methods, including super-resolution and noise removal, can enhance key rPPG-relevant features in both RGB and thermal imagery. This enhancement can lead to better recovery of physiological signals. Our comprehensive experiments on super-resolution and deblurring demonstrate that advanced deep learning models (e.g., SwinIR, EDSR, LapSRN) substantially improve the quality of RGB and thermal images, facilitating improved recovery of physiological signal/feature. These developments promote accessible, contactless cardiovascular care by providing the foundation for practical uses in telemedicine, home monitoring, workplace safety, and public health.

INTRODUCTION

Cardiovascular disease (CVD) remains the leading cause of morbidity and mortality globally, affecting millions of people and exerting a substantial burden on global healthcare systems[1]. Physiological health directly relies on heart condition and requires continuous monitoring to detect early warning signs and prevent adverse events, such as stroke and arrhythmias. Traditional health monitoring methods, including electrocardiography (ECG) and photoplethysmography (PPG), require specialized equipment, clinical visits, and direct skin contact. These methods are expensive and impractical for long-term use. It requires direct clinical professional involvement and can be uncomfortable due to direct skin contact. These limitations introduced an obstacle to continuous cardiac monitoring for individuals requiring long-term surveillance, particularly elderly patients and those in remote areas[2].

In response to these limitations, remote photoplethysmography (rPPG) has emerged as an innovative technique that can extract cardiovascular data without requiring physical contact. Remote photoplethysmography (rPPG) [3] is a non-invasive technique that uses subtle changes in blood volume to determine physiological signals, including pulse rate (PR), pulse rate variability (PRV), and respiratory rate. Compared to traditional techniques such as ECG and PPG [4], which require direct contact with the skin, rPPG offers a more practical and comfortable monitoring option, using remote cameras/sensors

to capture changes in light absorption and reflection caused by blood flow [5]. An increasing number of studies have been reported on remote rPPG in recent years, which utilize video or ordinary cameras to achieve non-contact remote detection of physiological signals[6].

The non-invasive nature of rPPG and seamless integration offer various practical applications, including healthcare, everyday life, and research. In the context of telemedicine, remote photoplethysmography (rPPG) enables medical professionals to assess patients' cardiovascular health during virtual consultations using a simple video call setup [7]. This functionality is particularly beneficial in emergency triage situations where rapid vital sign evaluation is critical, as well as in managing chronic illnesses and conducting routine checkups [8].

Another revolutionary application is home monitoring, which allows individuals to continuously check their cardiovascular health without needing any additional gadgets. While watching TV, using a computer, or going about their daily activities, people with chronic illnesses or elderly individuals living alone can all benefit from this seamless monitoring [9]. This ambient monitoring method not only provides doctors with comprehensive, long-term information about their patients' health but also removes barriers to compliance [10].

Incorporating rPPG technology in the workplace presents exclusive prospects for occupational health surveillance. The integration of rPPG technology in the workplace presents unprecedented potential for occupational health monitoring. It is possible to monitor workers in physically demanding jobs, high-stress workplaces, or safety-critical tasks for indications of exhaustion, stress, or cardiovascular discomfort without interfering with their work or forcing them to wear extra gear. This capability has the potential to transform workplace safety and wellness programs entirely [11].

Other potential use cases include applications related to security and public health. Airport screening systems could incorporate rPPG to identify individuals experiencing medical distress or elevated stress levels, enhancing biosecurity without physical contact [12]. Similarly, research applications in human-computer interaction, psychology, and behavioral studies hold another promising prospect. With rPPG, researchers can examine physiological correlates of different stimuli, stress reactions, and emotional responses in natural environments without the limitations imposed by wearable sensors. This potential could improve our knowledge of human behavior, emotion control, and the physiological effects of different settings or interventions [13].

In conclusion, remote rPPG effectively addresses significant gaps in cardiovascular disease management by enabling contactless and scalable monitoring, particularly important given the increasing prevalence of these health issues. With advancements in machine learning, this technology could revolutionize preventative healthcare, allowing the tracking of vital signs as simple as facing a camera.

1 PREPROCESSING FOR STREAM-NET

In the initial phase of the project, a dedicated preprocessing technique [Annex 1] was developed and implemented to prepare raw RGB face image sequences for input into the dual-branch deep learning model. These preprocessing steps are essential for ensuring that the spatio-temporal features extracted by the deep learning network are both accurate and reliable. This preprocessing pipeline is a vital part of the STREAM-Net framework, which was recently published in a Q1 journal [14].

The STREAM-Net model was trained and evaluated on the PURE [15] and UBFC [16] datasets, which are open-source RGB facial imaging data for physiological signal estimation. These datasets are collected under a controlled environment, including illumination and motion.

1.1 Key Preprocessing Steps

The preprocessing pipeline involved several steps, including image resizing, normalization, and face cropping. Below are details of these steps.

- **Input Resizing:** The raw RGB image sequences were resized to a fixed spatial resolution of 144×144 pixels. This uniformity in size reduces computational cost while ensuring consistent input shape across all samples.
- **Dual-Branch Input:** STREAM-Net is a dual-branch architecture consisting of a spatial and temporal branch, each receiving a different pre-processed input. The spatial branch takes a downsampled raw frame $c(t)$, while the temporal branch receives a normalized frame, calculated using the difference between two consecutive frames.

The normalized frames are calculated as:

$$\text{Normalized Frame } (t) = \frac{c(t+1) - c(t)}{c(t+1) + c(t)}$$

where $c(t)$ and $c(t+1)$ are consecutive raw frames in a sequence.

This normalization emphasizes dynamic skin changes (e.g., due to blood flow), while reducing the influence of static visual content like lighting or skin tone. Normalized features are less sensitive to tiny changes in skin tone or illumination,

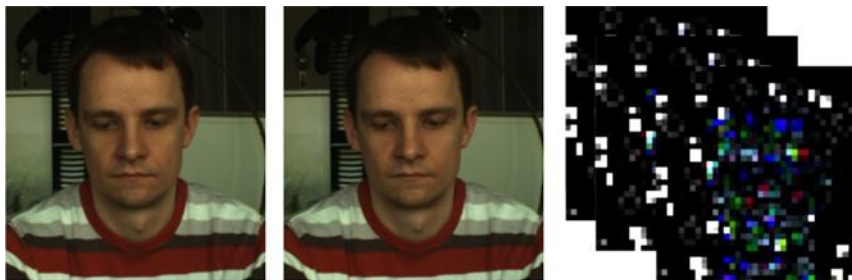


Figure 1: Raw frames from the PURE dataset and the normalized R, G, and B frames.

which promotes better generalization across datasets. A raw frame and a normalized frame of two consecutive frames are shown in Figure 1.

- **Face Cropping:** Before normalization, face cropping is applied to crop the region of interest (ROI) and discard background pixels that do not contribute to skin variation.

This preprocessing pipeline ensures the input to STREAM-Net is both spatially and temporally consistent, with reduced susceptibility to nuisance variables such as illumination and skin tone. It enhances the model's ability to extract robust features for downstream tasks like subtle variation on the skin surface due to blood pulsation.

2 MC-DROPOUT FOR UNCERTAINTY ESTIMATION

To quantify the uncertainty of the STREAM-Net, we employ the Monte Carlo dropout (MC-dropout) technique. The MC-Dropout method utilizes dropout during inference to estimate epistemic uncertainty. Dropout is a regularization technique used in deep neural networks to avoid overfitting problems. It randomly deactivates a subset of neurons from the learning process during the training phase. Figure 2 illustrates the role of dropout during the training and inference for both model configurations, including the standard model and the Monte Carlo dropout model for uncertainty estimation.

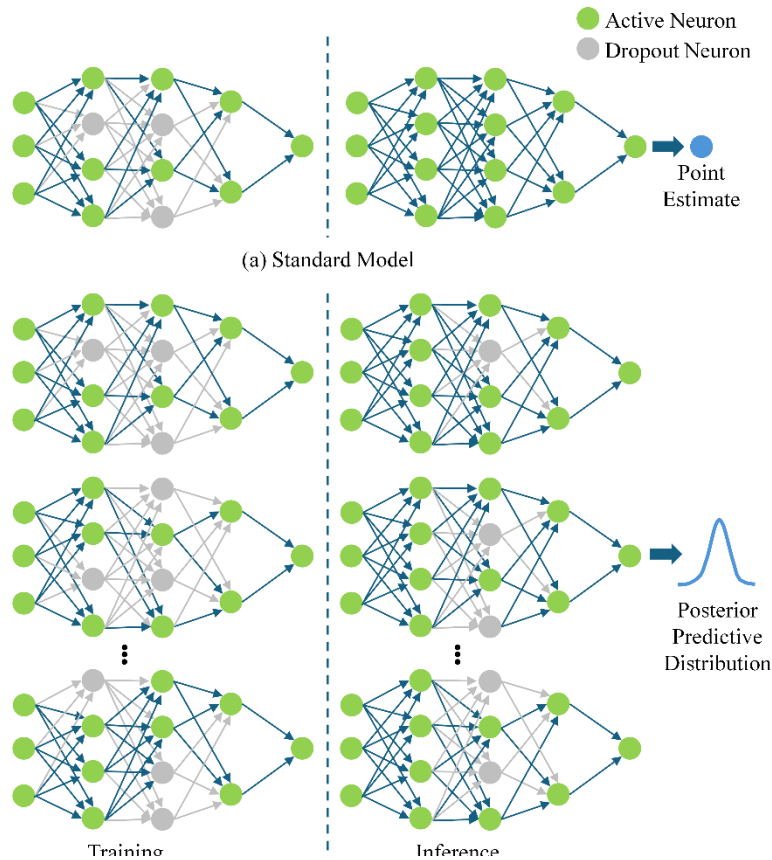


Figure 2: Model configurations for deterministic model prediction and Monte Carlo dropout for uncertainty quantification.

2.1 Implementation

To estimate the uncertainty of the STREAM-Net, we employed Monte Carlo (MC) dropout during the inference. This uncertainty analysis uses four different dropout configurations, including dropout rates of 0.1, 0.2, dropout div 2, and dropout div 3. Dropout div 2 and dropout div 3 denote that the original dropout used for training is divided by 2 and 3, respectively.

During the inference, MC dropout randomly enables the dropout layers and performs 30 predictions against a single input sample over 30 forward iterations.

2.2 MC-Dropout Predictions and Visualization

During the inference, MC dropout randomly enables the dropout layers and performs 30 predictions against a single input sample over 30 forward iterations. Tables 1 and 2 compare the quantitative analysis of the proposed standard model and the model with dropout enabled during inference. It is noticeable that the proposed model presents similar performance metrics using different dropout configurations, ensuring consistence performance.

Table 1: Cross-dataset results, trained on PURE and tested on UBFC. Best results are shown in bold, and second-best results are underlined

Method	MAE	MAPE	RMSE	ρ
PhyNet [18]	2.370	2.158	2.481	0.990
DeepPhys [19]	1.213	1.421	2.902	0.990
TS-CAN [20]	1.301	1.501	2.871	0.990
EfficientPhys-C [21]	2.131	2.350	3.000	0.940
PhysFormer [22]	1.440	1.660	3.770	0.980
SpikingPhys [23]	2.801	2.810	-	0.830
PhysMamba [24]	<u>0.970</u>	-	1.930	0.990
FactorizePhys [25]	0.770	<u>1.340</u>	3.291	0.991
STREAM-Net [14] (Standard Model)	1.151	1.305	<u>2.715</u>	0.991
Dropout 0.1	1.151	1.305	2.715	0.991
Dropout 0.2	1.151	1.305	2.715	0.991
Dropout/2	1.151	1.305	2.715	0.991
Dropout/3	1.151	1.305	2.715	0.991

Table 2: Cross-dataset results, trained on UBFC and tested on PURE. The best results are highlighted in bold, while the second-best results are underlined

Method	MAE	MAPE	RMSE	ρ
PhyNet [18]	8.061	13.671	19.710	0.613
DeepPhys [19]	5.541	5.320	18.511	0.661

TS-CAN [20]	3.690	<u>3.390</u>	13.801	0.820
EfficientPhys-C [21]	5.471	5.401	17.041	0.710
PhysFormer [22]	12.920	23.921	24.361	0.472
SpikingPhys [23]	3.831	-	<u>5.701</u>	0.830
PhysMamba [24]	1.20	-	5.990	0.970
STREAM-Net [14] (Standard Model)	<u>1.318</u>	1.384	5.314	0.997
Dropout 0.1	1.318	1.384	5.314	0.997
Dropout 0.2	1.318	1.384	5.314	0.997
Dropout/2	1.318	1.384	5.314	0.997
Dropout/3	1.318	1.384	5.314	0.997
Dropout 0.9	10.189	10.463	22.612	0.389

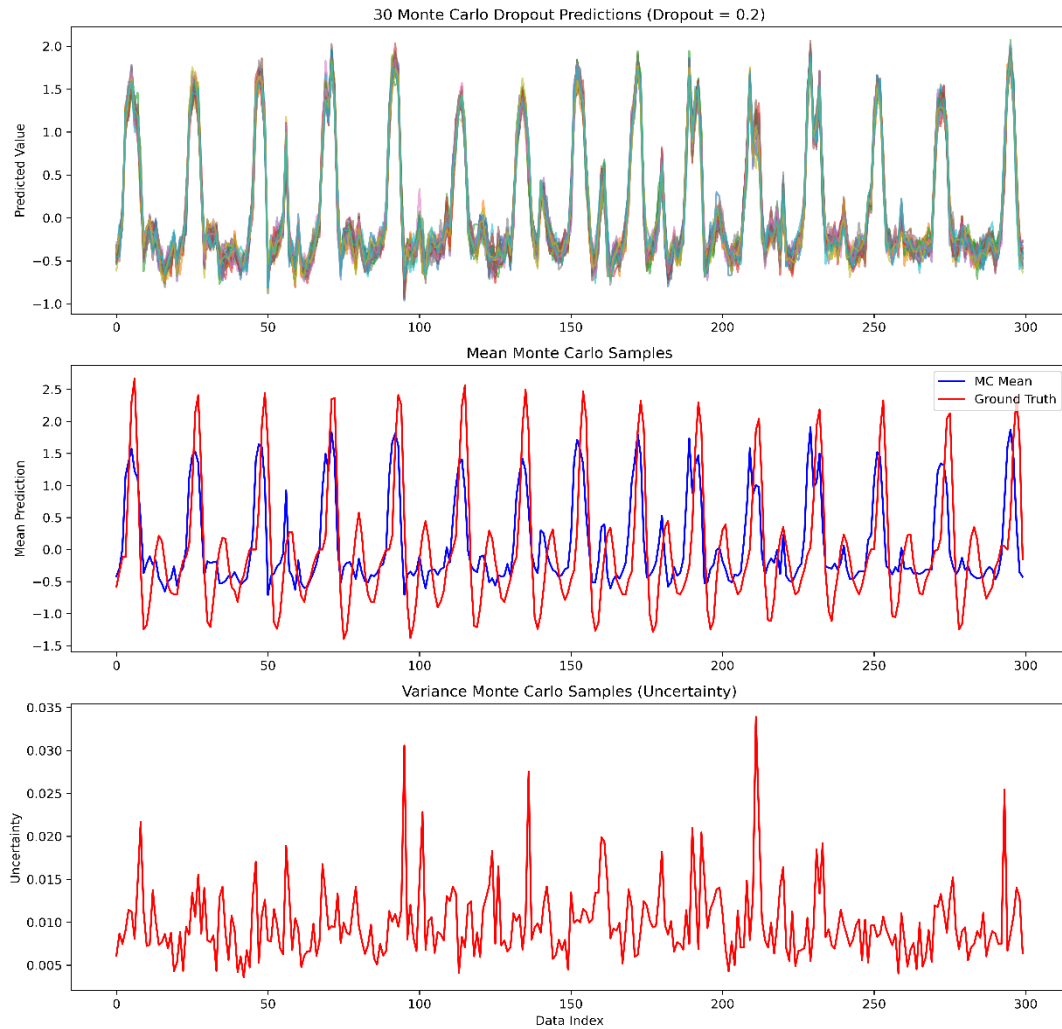


Figure 3: Monte Carlo dropout predictions and uncertainty analysis: The first plot presents 30 Monte Carlo (MC) dropout predictions (dropout rate = 0.2), illustrating the model's predictive uncertainty. The middle plot compares the mean MC prediction (blue) derived from MC, with the ground truth (red). The bottom plot depicts the variance across these predictions, quantifying the model's uncertainty over time steps (0-300).

Figure 3 presents a visual analysis of the uncertainty in the model predictions. The first plot shows 30 Monte Carlo dropout predictions (dropout rate = 0.2), obtained during 30 forward passes against one input sample. The middle plot compares the ground truth values (red) and the mean prediction (blue), obtained from Monte Carlo predictions. The variance of all 30 predictions is shown in the bottom plot, which measures the model's uncertainty over time steps (0-300). Similarly, Figure 4 illustrates the model's predictions under MC dropout with different dropout settings. A higher dropout rate (0.2) introduces higher prediction variance, while a lower dropout (Div 2 or Div 3) reduces the prediction variance and presents balanced stability and uncertainty [14].

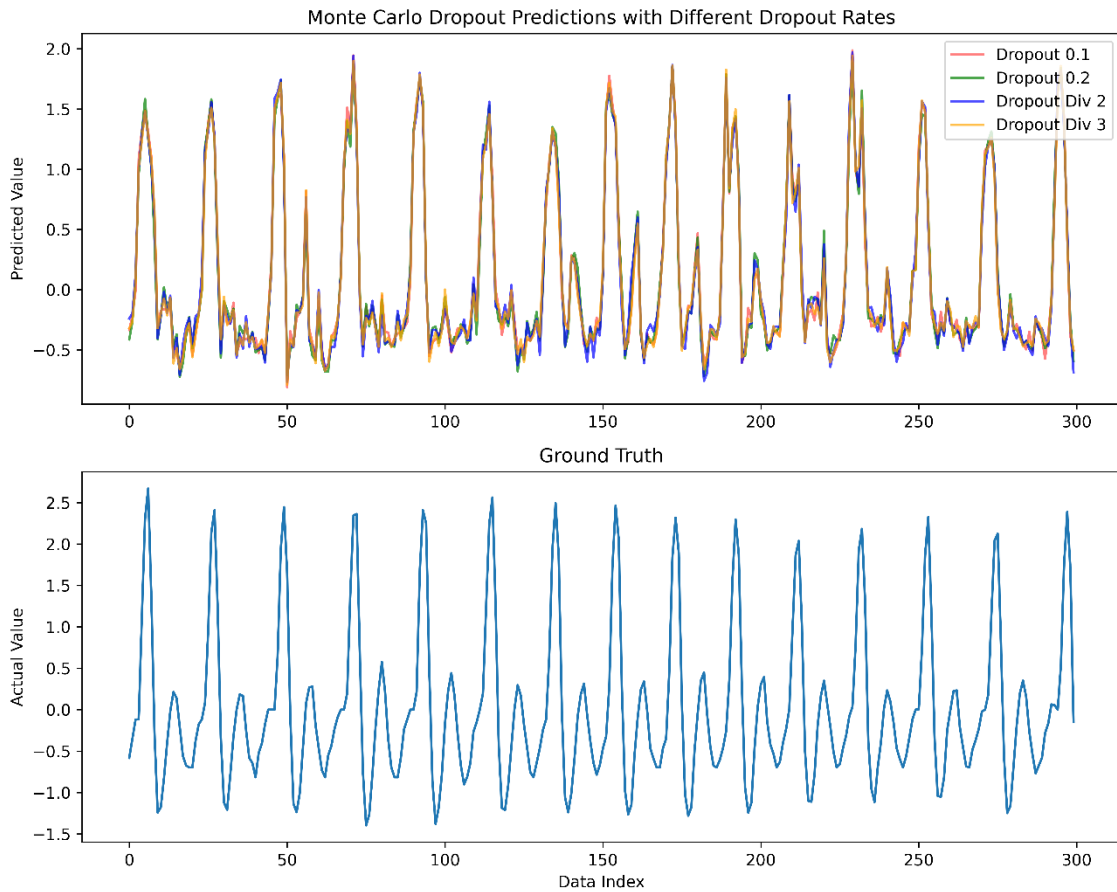


Figure 4: Monte Carlo dropout predictions uncertainty: Effect of varying dropout rates on model outputs relative to ground truth.

In general, the predictions remain consistent across all dropout rates, reflecting the model's ability to capture the underlying patterns of the data. The bottom subplot exhibits ground truth values for comparison. By comparing the MC dropout predictions with the ground truth, we can assess the model's ability to generalize while incorporating uncertainty estimation. This analysis highlights the model's robustness against different dropout rates and offers a method to quantify uncertainty in Bayesian deep learning

scenarios. It is interesting to observe that there is no difference between the predictions of the standard model (no dropout during inference) and the predictions under different dropout rate configurations. This consistence performance supports the reliability of our model under stochastic conditions. As these performance metrics are indirectly computed via the Fast Fourier Transform (FFT), small variations in the predicted values of the signal due to different dropout rates are insignificant and the overall performance remains unaffected. However, to ensure our uncertainty estimation algorithm, we applied extreme conditions by setting the dropout rate to 0.9. We observed a remarkable decline in predictive accuracy when using these extreme dropout settings (Table 2, dropout 0.9). This experimental analysis ensures that MC-dropout works properly and STREAM-Net exhibits consistent performance across different dropout rates.

In addition, Fig. 3 shows a visualization of these probabilistic predictions. The first plot in the figure depicts 30 forward passes on an input sample and the corresponding MC dropout prediction (with a dropout rate of 0.2). As can be seen from the figure, there are small variations in each of the predictions, but they are all consistent with the expected results, and there are no uncertain estimations. The second plot compared the mean of these probabilistic predictions with the ground truth. This implies that the MC-prediction

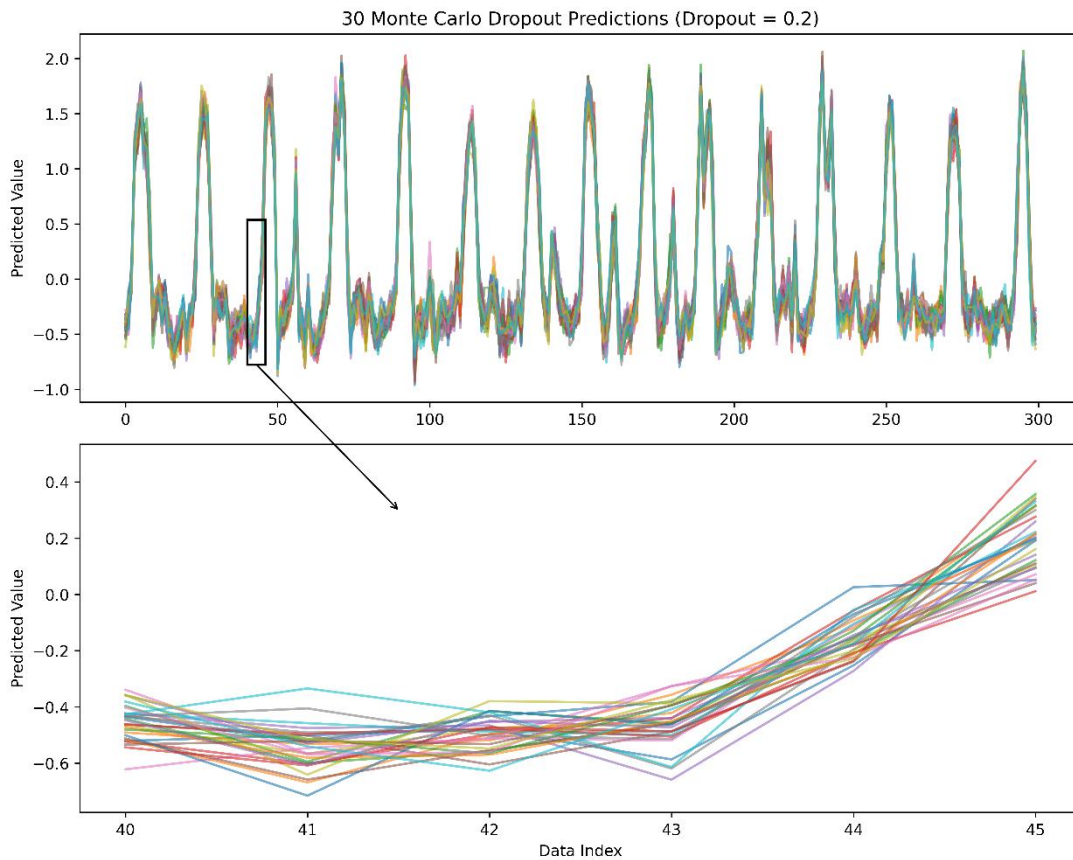


Figure 5: MC dropout predictions: A magnified view of the highlighted region (index 40 to 45) illustrating the variability in prediction results across different samples.

follows the data trend as the actual signal (ground truth). The last plot in Fig. 3 shows the uncertainty during these 30 inferences with randomly activated dropout layers. In general, the model uncertainty is low (0.05–0.015), indicating high confidence in the predictions. However, there are some instances where model exhibits higher uncertainty. For instance, the MC-dropout variance for data indices 100 and 210 is as high as 0.030. The complexity of the input data or the variability of the model predictions may lead to such a high uncertainty region. These fluctuations and spikes in uncertainty indicate that the model encounters varying degrees of uncertainty throughout the input data sequence.

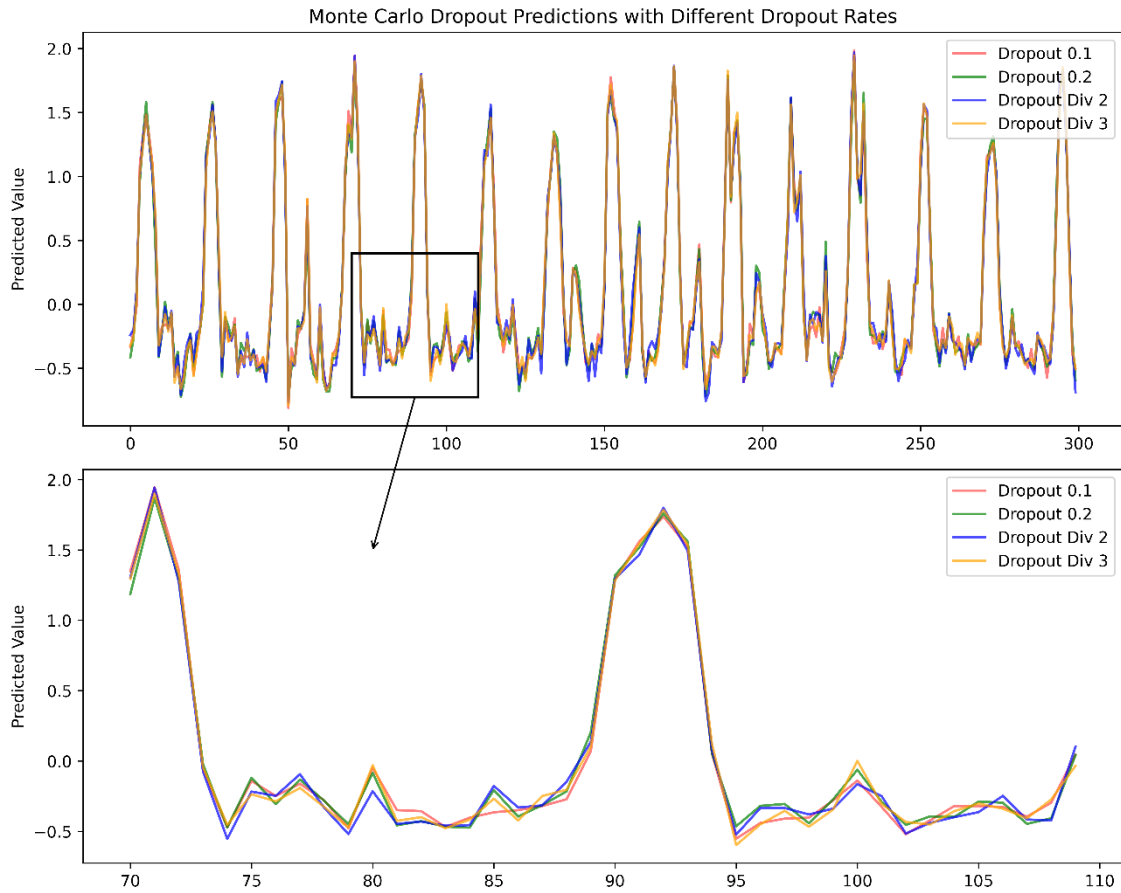


Figure 6: MC dropout rate predictions: A zoomed-in view of the highlighted region (indices 70 to 110) comparing the variability in predictions for different dropout rates.

To further understand the variability introduced by the MC-Dropout, Figs. 5 and 6 provide a closer view of the MC-Dropout predictions, illustrating the predicted variability associated with the MC-Dropout over a smaller range.

Table 3 provides the mean signal-to-noise ratio (SNR), mean variance and mean standard deviation of uncertainty to demonstrate the quantitative analysis using different dropout rates. These statistics reflect that higher the dropout rate, higher the uncertainty, and vice versa. For example, the mean, variance, and standard deviation for a dropout rate of 0.1

are 0.0049 and 0.002, respectively, whereas the higher the dropout rate (dropout = 0.2), the greater the uncertainty (mean variance = 0.0113, standard deviation = 0.0045). Overall, there is no significant difference in the uncertainty comparisons using different dropout settings compared to the standard model predictions. This validates the robust performance and adaptability of STREAM-Net under different input conditions. More details in published paper [14] and in Annex 1 (GitHub repository).

The MC-dropout experiments validates that STREAM-Net maintains consistent performance across different dropout configurations, with only extreme dropout (0.9) causing a decline in accuracy. This demonstrates the model's robustness, reliability under stochastic conditions, and its ability to provide meaningful uncertainty estimates for probabilistic predictions.

Table 3: Monte Carlo mean variance and standard deviation (Std) for uncertainty analysis.

Method	Mean SNR	Mean Variance	Variance Std.
No Dropout	5.739	-	-
Dropout 0.1	5.775	0.0049	0.0022
Dropout 0.2	5.812	0.0113	0.0045
Dropout/2	5.800	0.0113	0.0037
Dropout/3	5.768	0.0064	0.0021

3 SUPER-RESOLUTION (SR) ALGORITHMS

The objective of developing new super-resolution and deblurring deep learning (DL) algorithms is to enhance high-frequency details in thermal images and correct color distortions, key requirements for local photoplethysmography (PPG) and lesion differentiation. These experiments serve as a fundamental step towards high-quality multimodal data enhancement with high-frequency feature recovery and color correction for local PPG.

3.1 Experimental details

Input Data:

- A subset of facial RGB and thermal images from the BP4D dataset [17] was used to conduct experiments related to image resolution enhancement.
- Images were artificially subsampled by a factor of 2× and 4× using nearest neighbour (NN) interpolation to generate low-resolution images.

Evaluation Metrics:

To assess the reconstruction quality of the enhanced/upsampled images, the following two metrics were used.

- **PSNR (Peak Signal-to-Noise Ratio):**

PSNR quantifies the ratio between the maximum possible pixel intensity and the mean square error (MSE) between the reference image I (input image) and the reconstructed image $\hat{I}$ (super-resolution). It is quantified in decibels (dB). The higher the PSNR, the better the image fidelity.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right)$$

where MAX is the maximum possible pixel value of the image, and $MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - \hat{I}(i, j)]^2$.

- **SSIM (Structural Similarity Index):**

SSIM measures the perceptual similarity between two images, considering luminance, contrast, and structure. It ranges from -1 to 1, where 1 indicates perfect similarity.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the average luminance of the two images being compared. σ_x^2 and σ_y^2 are the variances of the pixel intensities in the two images. σ_{xy} is the covariance of the pixel intensities between the two images. C_1 and C_2 are constants added to prevent instability when the denominators are close to zero

3.2 Methods Evaluated

To perform the super-resolution experiments, we started from the classical interpolation methods for image enhancement, which include nearest neighbor and bilinear interpolation. We further extend our experiments with more advanced approaches, including OpenCV and a deep learning-based method for image reconstruction.

Baseline Methods:

- Nearest Neighbour Upsampling
- Bilinear Upsampling

These two methods serve as baseline techniques for image super-resolution and are fast but limited in reconstruction fidelity.

OpenCV DNN:

- EDSR (Enhanced Deep Residual Network)
- LapSRN (Laplacian Pyramid Super-Resolution Network)

Advance Deep Learning Methods:

We employed various deep learning-based super-resolution methods using our custom settings, including the recent method.

- **Swin IR**
- **Swin IR_L**
- **Struct SR**
- **Deep Fourier SR**
- **Face SR**

3.3 Results and Evaluation

We evaluated the above-mentioned image enhancement method using two performance measuring metrics: PSNR and SSIM. Table 4 compares the performance of these methods.

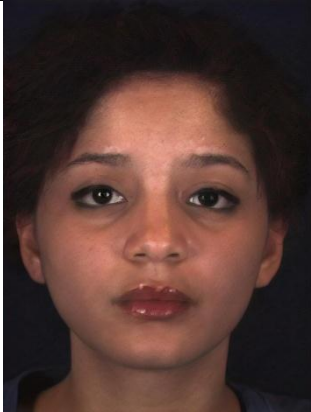
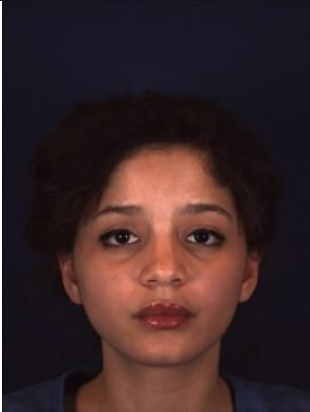
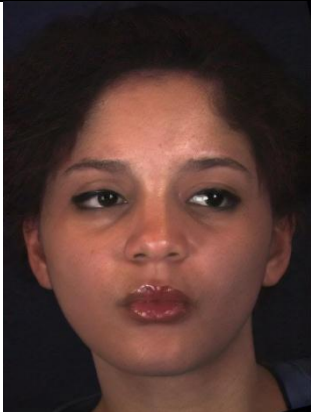
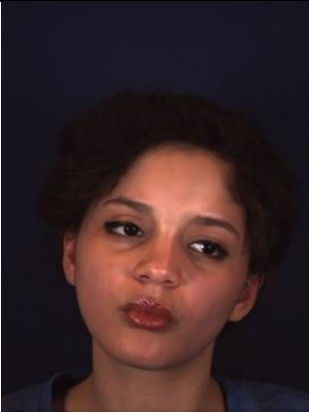
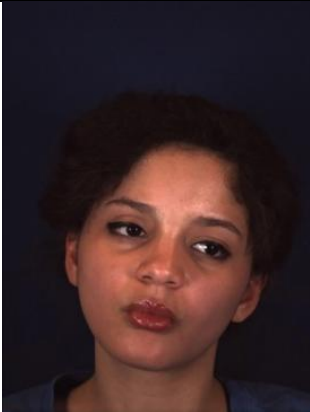
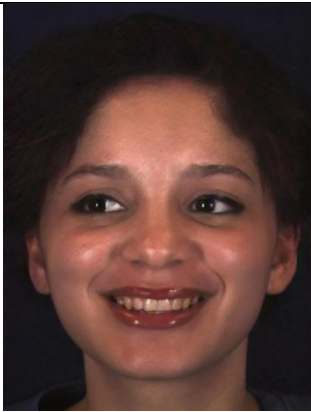
Table 4: Performance Comparison of Various Super-Resolution Methods

Method	PSNR (2x)	SSIM(2x)	PSNR(4x)	SSIM(4x)
NN	39.348	0.958	34.167	0.904
Bilinear	40.692	0.968	35.436	0.931
ESDR	-	-	35.401	0.931
LapSRN	-	-	39.024	0.947
SwinIR	13.00	0.666	34.63	0.915
SwinIR_L	-	-	35.39	0.933
StructSR	-	-	19.08	0.763
DeepFourierSR	12.39	0.579	12.42	0.569
FaceSR	-	-	15.32	0.7074

Qualitative Results

To further evaluate the performance of image enhancement methods, we compare the qualitative analysis in Table 5. This qualitative comparison presents visual comparisons between the low-resolution input images (1x) and the reconstructed high-resolution images (4x) using different deep learning algorithms including, FaceSR, EDSR, and LapSRN. We choose three different samples of images with low resolution and input to three different super-resolution enhancements methods to reconstruct high resolution images. In contrast the low resolution input images, reconstructed images presents enhanced perceptual quality and higher resolution, showing capacity for fine-detail signal recovery. To conclude, these results demonstrate that deep learning-based super-resolution methods can substantially improve image quality, thereby facilitating improved feature extraction for more accurate rPPG signal estimation.

Table 5: Qualitative comparison of low-resolution image (1x) and reconstructed super-resolution image (4x). These images are taken from BP4D dataset [17].

Low Resolution Image (1x)	Processed High Resolution Image (4x)		
	FaceSR	ESDR	LapSRN
			
			
			

The experimental results demonstrate that while classical interpolation methods provide reasonable baseline performance, advanced deep learning-based approaches, particularly LapSRN and SwinIR variants, achieve superior reconstruction quality in terms of both PSNR and SSIM. Qualitative comparisons further confirm that these methods recover finer details and improve perceptual quality, which is essential for accurate local PPG signal estimation and lesion differentiation. Overall, super-resolution and deblurring with deep learning form a crucial step towards high-quality multimodal data enhancement.

4 FUTURE WORK FOR D2.6

The super-resolution experiments reported in D2.1 form the foundation for enhancing image quality in the multimodal database. In deliverable D2.6, this work will be extended to develop a comprehensive Multimodal Database of Face and Skin Vascularization Maps (M40). Future work will focus on robust rPPG estimation using multimodal imaging, combining thermal, and RGB image data. This includes multimodal data collection, advanced preprocessing techniques for image registration, noise filtration and feature enhancement. Furthermore, design a custom DL model to capture both spatial and temporal physiological signals using collected multimodal data. Together, these efforts aim to enable accurate rPPG estimation from enhanced multimodal images, enabling remote health monitoring such as rPPG signal estimation and facial skin analysis using vascular maps.

REFERENCES

1. Chong, Bryan et al. "Global burden of cardiovascular diseases: projections from 2025 to 2050." *European journal of preventive cardiology*, zwae281. 13 Sep. 2024, doi:10.1093/eurjpc/zwae281
2. Debnath, U., & Kim, S. (2025). A comprehensive review of heart rate measurement using remote photoplethysmography and deep learning. *BioMedical Engineering OnLine*, 24(1), 73.
3. Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics express*, 16(26), 21434-21445.
4. Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of medical engineering & technology*, 33(8), 634-641.
5. Daly, S. M., & Leahy, M. J. (2013). 'Go with the flow': a review of methods and advancements in blood flow imaging. *Journal of biophotonics*, 6(3), 217-255.
6. Ni, A., Azarang, A., & Kehtarnavaz, N. (2021). A review of deep learning-based contactless heart rate measurement methods. *Sensors*, 21(11), 3719.
7. Khanam, F. T. Z., Al-Naji, A., & Chahl, J. (2019). Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: A review. *Applied Sciences*, 9(20), 4474.
8. Pirzada, P., Wilde, A., Doherty, G. H., & Harris-Birtill, D. (2023). Remote photoplethysmography (rPPG): A state-of-the-art review. *MedRxiv*, 2023-10.

9. Caroppo, A., Manni, A., Rescio, G., Siciliano, P., & Leone, A. (2024). Vital signs estimation in elderly using camera-based photoplethysmography. *Multimedia Tools and Applications*, 83(24), 65363-65386.
10. Cristina, S., Počta, P., Zgank, A., Camilleri, K. P., Colantonio, S., & Lambrinos, L. (2025). Remote monitoring of vital signs. In *Privacy-Aware Monitoring for Assisted Living: Ethical, Legal, and Technological Aspects of Audio-and Video-Based AAL Solutions* (pp. 217-237). Cham: Springer Nature Switzerland.
11. Ahmed, S. G., Verbert, K., Siedahmed, N., Khalil, A., Aljassmi, H., & Alnajjar, F. (2025). AI Innovations in rPPG Systems for Driver Monitoring: Comprehensive Systematic Review and Future Prospects. *Ieee Access*.
12. Adisa, O., Akinduyite, O., & Badeji-Ajisafe, B. (2024, November). A Comprehensive Review of Remote Photoplethysmography Techniques for Accurate and Robust Heart Rate Monitoring. In *2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)* (pp. 1-5). IEEE.
13. Fontes, L., Machado, P., Vinkemeier, D., Yahaya, S., Bird, J. J., & Ihianle, I. K. (2024). Enhancing stress detection: A comprehensive approach through rPPG analysis and deep learning techniques. *Sensors*, 24(4), 1096.
14. Usman, M., Sobotka, M., & Ruminski, J. (2025). STREAM-Net: Spatio-Temporal Feature Fusion Network for Robust rPPG Signal Measurement in Remote Health Monitoring. *Knowledge-Based Systems*, 114080.
15. R. Stricker, S. Müller, H.-M. Gross, Non-contact video-based pulse rate measurement on a mobile service robot, in: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE, 2014, pp. 1056–1062.
16. Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., & Dubois, J. (2019). Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern recognition letters*, 124, 82-90.
17. Zhang, Zheng, Jeff M. Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan et al. "Multimodal spontaneous emotion corpus for human behavior analysis." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3438-3446. 2016.
18. Yu, Zitong, Xiaobai Li, and Guoying Zhao. "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks." *arXiv preprint arXiv:1905.02419* (2019).
19. Chen, Weixuan, and Daniel McDuff. "Deepphys: Video-based physiological measurement using convolutional attention networks." In *Proceedings of the european conference on computer vision (ECCV)*, pp. 349-365. 2018.
20. Liu, X., Fromm, J., Patel, S., & McDuff, D. (2020). Multi-task temporal shift attention networks for on-device contactless vitals measurement. *Advances in Neural Information Processing Systems*, 33, 19400-19411.
21. Liu, X., Hill, B., Jiang, Z., Patel, S., & McDuff, D. (2023). Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 5008-5017).
22. Yu, Z., Shen, Y., Shi, J., Zhao, H., Torr, P. H., & Zhao, G. (2022). Physformer: Facial video-based physiological measurement with temporal difference

- transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4186-4196).
23. Liu, M., Tang, J., Chen, Y., Li, H., Qi, J., Li, S., ... & Chen, H. (2025). Spiking-PhysFormer: Camera-based remote photoplethysmography with parallel spike-driven transformer. *Neural Networks*, 185, 107128.
 24. Luo, C., Xie, Y., & Yu, Z. (2024, November). PhysMamba: Efficient remote physiological measurement with SlowFast temporal difference mamba. In *Chinese Conference on Biometric Recognition* (pp. 248-259). Singapore: Springer Nature Singapore.
 25. Joshi, J., Agaian, S. S., & Cho, Y. (2024). Factorizephys: Matrix factorization for multidimensional attention in remote physiological sensing. *arXiv preprint arXiv:2411.01542*.

ANNEX

1. <https://github.com/usmanraza121/STREAM-Net>