Bachelor of Science in Electrical and Electronic Engineering
EEE 400 (January 2024): Thesis

Heart Rate Estimation From PPG Signal Using Contrastive Learning

Submitted by

Ameer Hamja Ibne Jamal
201906177

Zahin Tazwar
201906183

Supervised by

Dr. Hafiz Imtiaz
Professor

Department of Electrical and Electronic Engineering

Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

March 2025

CANDIDATES' DECLARATION

This is to certify that the work presented in this thesis, titled, "Heart Rate Estimation From PPG Signal Using Contrastive Learning", is the outcome of the investigation and research carried out by us under the supervision of Dr. Hafiz Imtiaz.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Ameer Hamja Ibne Jamal
201906177

Zahin Tazwar
201906183

CERTIFICATION

This thesis titled, **"Heart Rate Estimation From PPG Signal Using Contrastive Learning"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for EEE 400: Project/Thesis course, and as the requirements for the degree B.Sc. in Electrical and Electronic Engineering in March 2025.

Group Members:

Ameer Hamja Ibne Jamal

Zahin Tazwar

Supervisor:

Dr. Hafiz Imtiaz
Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology

ACKNOWLEDGEMENT

We are indebted to all those who have contributed to the successful completion of this research.

First and foremost, we express our deepest gratitude to our supervisor, Dr. Hafiz Imtiaz, for his invaluable guidance, encouragement, and continuous support throughout this study. His expertise, insightful feedback, and patience have played a crucial role in shaping the direction of our research and refining our understanding of the subject matter.

We are also profoundly grateful to our families for their unwavering support, patience, and encouragement. Their belief in our abilities and constant motivation have been a driving force behind our perseverance in this academic journey.

Additionally, we extend our sincere appreciation to our roommates and friends for their steadfast support, insightful discussions, and encouragement during challenging moments. Their presence has been invaluable in keeping us motivated and focused on our goals.

Finally, we acknowledge everyone who has contributed, directly or indirectly, to this research. Your support and assistance have been instrumental in making this work possible, and we are truly grateful.

Dhaka
March 2025

Ameer Hamja Ibne Jamal

Zahin Tazwar

Contents

List of Figures

List of Tables

ABSTRACT

Accurate heart rate monitoring is essential for effective health assessment, enabling early detection of cardiovascular abnormalities and facilitating personalized healthcare interventions. While electrocardiography (ECG) is the most reliable method for heart rate detection, providing high-fidelity measurements of cardiac electrical activity, its implementation can be challenging in many real-world scenarios. This is due to the complexity of data acquisition, which often requires specialized equipment, trained personnel, and direct skin contact, limiting its practicality for continuous and unobtrusive monitoring. Photoplethysmography (PPG) offers a more accessible alternative for heart rate monitoring, utilizing optical sensors to measure blood volume changes in peripheral tissues. However, its reliability is often compromised by various factors that introduce noise and artifacts into the signal, such as motion, ambient light interference, and variations in skin perfusion. These limitations hinder the widespread adoption of PPG-based heart rate monitoring in clinical and consumer applications.

To address the limitations of PPG and improve the accuracy of heart rate estimation, we propose a novel contrastive learning approach. Contrastive learning enables the model to learn discriminative feature representations by comparing positive and negative sample pairs, enhancing robustness to noise and variability. Our method utilizes a ConvNeXt architecture, a modern convolutional neural network known for its efficiency and effectiveness in feature extraction, to extract robust feature representations from both PPG and ECG signals. By leveraging ECG signals as a reference, we provide additional information to guide the learning process and improve the accuracy of PPG-based heart rate estimation. These representations are then integrated to form a combined embedding space, capturing relevant information from both modalities. A Rank-N-Contrast (RnC) loss function is employed to optimize the model, specifically designed for regression tasks by focusing on the relative order of heart rate values. We evaluated our approach using the PPG-DaLiA dataset, a publicly available benchmark dataset for heart rate estimation, and achieved a mean absolute error of **5.84** beats per minute. This result demonstrates the efficacy of the proposed method for accurate heart rate estimation from PPG, showing a significant improvement over traditional PPG-based methods and highlighting its potential for enhancing wearable health monitoring technologies.

Keywords: Contrastive Learning, PPG, ECG, Heart Rate Dete.

Chapter 1

Introduction

1.1 Background

Heart rate estimation plays a crucial role in monitoring cardiovascular health, fitness, and overall well-being. Traditionally, electrocardiograms (ECG) have been the gold standard for heart rate measurement, but they require physical contact with the skin and are often bulky and expensive. An alternative, more practical approach is photoplethysmogram (PPG), a non-invasive and low-cost method that uses light absorption to measure blood volume changes in the microvascular bed of tissue. PPG signals can be easily captured through wearable devices like smartwatches and fitness trackers, making them suitable for continuous, real-time heart rate monitoring. Despite its advantages, PPG-based heart rate estimation faces several challenges, such as noise, motion artifacts, and the non-stationary nature of the signal.

Recent advancements in machine learning, particularly deep learning, have significantly improved the accuracy of heart rate estimation from PPG signals. These approaches have moved beyond traditional signal processing techniques, enabling automatic feature extraction and robust regression models for heart rate prediction. However, the development of such systems requires overcoming several obstacles, such as the limited availability of labeled data and the presence of motion artifacts that distort PPG signals.

1.2 Motivation

The motivation behind this thesis is to explore the application of representation learning techniques, specifically contrastive learning, to improve heart rate estimation from PPG data. Contrastive learning, a powerful self-supervised learning method, allows the model to learn discriminative features from unlabeled data by comparing positive and negative samples. This approach is particularly beneficial when labeled data is scarce, a common issue in medical and healthcare-related domains [1, 2].

Furthermore, the thesis investigates the use of deep learning architectures, such as Convolutional Neural Networks (CNNs) and Bi-directional Long Short-Term Memory networks (BiLSTM), in combination with contrastive learning to enhance feature extraction and temporal sequence modeling. While traditional architectures have been explored for PPG-based heart rate estimation, the ConvNext architecture, a more recent and efficient deep learning model, is introduced to address these challenges [3]. The performance of the proposed model is evaluated and compared to state-of-the-art methods like KID-PPG and Deep PPG [4, 5].

1.3 Problem Statement

Heart rate estimation from PPG data is prone to errors due to various factors, such as noise, motion artifacts, and variations in signal quality. Traditional approaches rely heavily on hand-crafted features, which may not fully capture the complexities of PPG signals. Additionally, while deep learning models have shown promise, the challenge remains to develop an efficient method that can handle both the large variability in PPG signals and the scarcity of labeled data. The aim of this research is to improve heart rate estimation accuracy by leveraging contrastive learning and state-of-the-art deep learning architectures.

1.4 Research Objectives

The main objectives of this thesis are:

- To investigate the use of contrastive representation learning for heart rate estimation from PPG signals.
- To develop a robust model using the Rank-N-Contrast Loss for heart rate regression tasks [6].
- To explore the effectiveness of the ConvNext architecture for feature extraction and heart rate prediction from PPG data [3].
- To compare the performance of the proposed model with existing methods like KID-PPG and Deep PPG [4, 5].
- To explore techniques for removing motion artifacts from PPG data to improve estimation accuracy [7].

1.5 Contributions of the Thesis

This thesis makes the following contributions:

- **Application of Rank-N-Contrast Loss:** A novel contrastive loss function, tailored for regression tasks, that enhances the model's ability to learn discriminative features from PPG signals [6].
- **Comparison with State-of-the-art Models:** The proposed model is evaluated against existing methods such as KID-PPG and Deep PPG to benchmark its performance [4, 5].
- **Application of ConvNext Architecture:** The ConvNext model is used to extract meaningful features from PPG signals, offering improved accuracy compared to traditional CNN architectures [3].

1.6 Thesis Structure

This thesis is organized as follows:

- **Chapter 2: Literature Review:** Provides a comprehensive review of the relevant literature on PPG-based heart rate estimation, contrastive learning, deep learning architectures, and motion artifact removal techniques.
- **Chapter 3: Methodology:** Describes the methodology used in this research, including the proposed model architecture, data preprocessing techniques, and the experimental setup.
- **Chapter 4: Results and Discussion:** Presents the results of the experiments and compares the proposed model's performance with state-of-the-art methods.
- **Chapter 5: Conclusion and Future Work:** Summarizes the findings of the thesis and discusses potential future directions for improving heart rate estimation from PPG data.

This chapter provided an overview of heart rate estimation using PPG data, highlighting the challenges posed by noise and motion artifacts. It outlined the motivation for using contrastive learning in conjunction with deep learning architectures to improve the accuracy of heart rate estimation. The next chapter will review existing research and methods in this domain, setting the foundation for the proposed approach.

Chapter 2

Literature Review

2.1 Introduction

Heart rate (HR) monitoring is essential for cardiovascular health assessment, fitness tracking, and stress analysis. PPG signals, which measure blood volume changes in peripheral vessels, provide a cost-effective and convenient method for HR estimation [8]. However, real-world PPG signals often suffer from motion artifacts, ambient light interference, and sensor variations, making traditional signal processing methods inadequate.

Deep learning has recently been applied to improve HR estimation from PPG signals. By leveraging feature extraction capabilities of neural networks, these models can enhance signal quality, filter noise, and improve HR accuracy [9]. This review examines deep learning-based HR estimation methods, including CNNs, RNNs, and hybrid models, while also discussing dataset availability and key challenges in the field.

2.2 Deep Learning Approaches for HR Estimation

Several deep learning models have been proposed for improving HR estimation accuracy from PPG signals.

2.2.1 CNN-based Approaches

CNNs are effective in extracting spatial and spectral features from PPG signals. These models process time-series data by applying 1D and 2D convolutional filters, which capture periodic patterns associated with heart rate [10]. Studies have demonstrated that CNN-based models outperform traditional HR estimation techniques, especially in dealing with motion artifacts

[11].

2.2.2 RNN and LSTM Models

RNNs, particularly LSTMs, are widely used for sequential data processing. These models are capable of capturing temporal dependencies in PPG signals, leading to improved HR estimation accuracy [12]. LSTMs have been successfully applied to process raw PPG signals and predict HR values in real-time applications [13].

2.2.3 Hybrid CNN-LSTM Models

Hybrid models combining CNNs and LSTMs have shown superior performance in HR estimation tasks. CNNs extract spatial features from PPG signals, while LSTMs capture long-term temporal dependencies, resulting in more robust HR estimation [14]. Some studies have explored transformer-based models to further enhance prediction accuracy [15].

2.3 Public Datasets for HR Estimation

Several publicly available datasets have facilitated the development of deep learning-based HR estimation models:

- **IEEE PPG-DaLiA** [8] – Contains PPG and ECG signals recorded under various physical activities.
- **MIMIC-III** [11] – A comprehensive database of physiological signals, including PPG, ECG, and HR recordings.
- **UBFC-RPPG** [12] – A dataset designed for remote photoplethysmography (rPPG) studies.
- **CapnoBase** [15] – A dataset including synchronized PPG and respiratory signals.

These datasets play a crucial role in training and validating deep learning models for HR estimation.

2.4 Challenges in HR Estimation from PPG

Despite significant advancements, several challenges remain:

2.4.1 Motion Artifacts and Noise

PPG signals are highly susceptible to motion artifacts, which introduce noise and distort HR estimations. Deep learning models have been developed to mitigate these effects by leveraging adversarial training and denoising techniques [9].

2.4.2 Generalization Across Different Populations

Most deep learning models are trained on specific datasets and may not generalize well to diverse populations with varying skin tones, physiological conditions, and sensor types [10].

2.4.3 Computational Complexity and Real-Time Processing

Deploying deep learning models on wearable devices requires optimization techniques such as model quantization, pruning, and knowledge distillation to achieve real-time performance [14].

2.5 Future Directions

Future research should focus on:

- Developing lightweight deep learning models for wearable devices.
- Improving dataset diversity to enhance generalization across different populations.
- Integrating multimodal data (ECG, accelerometer, and PPG) for improved HR estimation.
- Exploring transformer-based models for better sequential data processing.

2.6 Conclusion

Deep learning has significantly improved HR estimation from PPG signals by enhancing signal quality and robustness against noise. However, challenges such as motion artifacts, dataset biases, and real-time implementation must be addressed. Future research should focus on optimizing deep learning models for real-world applications, ensuring scalability and generalization.

Chapter 3

Methodology

3.1 Introduction

Heart rate estimation from photoplethysmogram (PPG) signals is a critical problem in healthcare applications. In this work, we explore a novel approach that integrates contrastive learning with deep neural networks to improve the accuracy of heart rate estimation from PPG signals. The existing methods primarily utilize regression-based models, signal processing techniques, and handcrafted features, which often struggle to account for noise and motion artifacts inherent in PPG data. The primary goal of this research is to investigate whether contrastive learning, which has been successful in other domains like image recognition, can enhance feature extraction and generalization in the context of PPG-based heart rate estimation.

The methodology proposed in this thesis includes the use of **contrastive learning** through **Rank-N-Contrast Loss (RnC)**, specifically designed for regression tasks. This is in contrast to traditional contrastive learning techniques such as SimCLR [1] and SupCon [2], which are typically used for classification tasks. For feature extraction, we employ the **ConvNeXt** architecture [3], a state-of-the-art model in convolutional neural networks (CNNs) that integrates transformer-like features to capture both short-term and long-term dependencies in time-series data like PPG signals.

In this chapter, we detail the key components of the methodology, including the problem formulation, the proposed model architecture, loss functions, training procedures, and data preprocessing steps. We also provide a rigorous mathematical formulation to represent the heart rate estimation process, ensuring clarity and precision in the description of each methodology component.

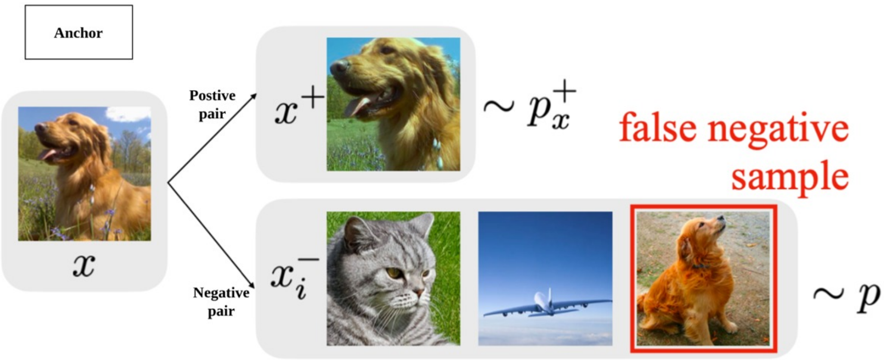

Figure 3.1: A visual representation of contrastive learning, demonstrating the formation of positive and negative pairs from an anchor image.

3.2 Problem Formulation

The problem of heart rate estimation from PPG signals is framed as a regression problem. Given a time series of PPG signals $\mathbf{x} = (x_1, x_2, \ldots, x_T)$, where T is the number of time steps, the objective is to predict the heart rate $y \in R$. The regression model, therefore, learns a mapping f from the input signal $\mathbf{x}$ to the continuous heart rate value y.

$$y = f(\mathbf{x}; \theta)$$

where $f(\cdot)$ is a model function parameterized by θ, and y is the output heart rate value corresponding to the input sequence $\mathbf{x}$.

Since the PPG signals are subject to noise and motion artifacts, the model must be robust enough to handle these distortions. Therefore, learning an effective representation of the signal is key to improving the model's ability to generalize well across unseen data.

3.3 Contrastive Learning for Heart Rate Estimation

Traditional contrastive learning methods like SimCLR [1] and SupCon [2] were primarily developed for classification tasks. These methods learn a feature embedding by contrasting pairs of positive and negative samples, where the positive pairs are similar (e.g., images of the same object) and the negative pairs are dissimilar (e.g., images of different objects). While this approach is well-suited for classification tasks, heart rate estimation is a regression problem. As a result, the direct application of traditional contrastive learning methods does not provide the desired improvements for regression tasks.

To address this limitation, we adopt **Rank-N-Contrast (RnC) Loss** [6], which is specifically

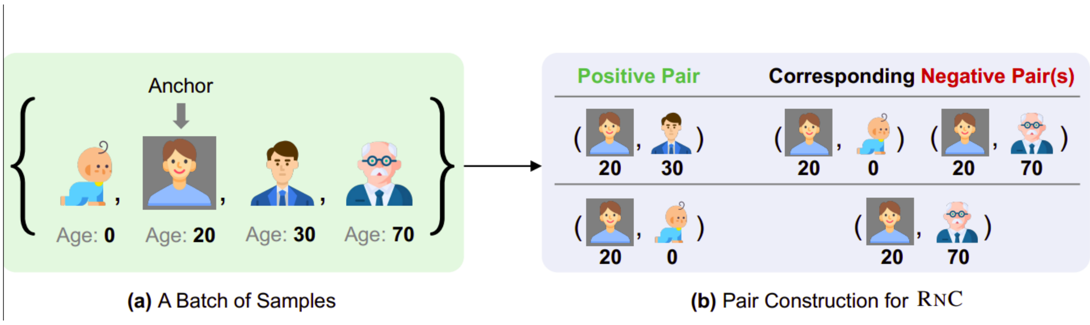

Figure 3.2: Illustration of Rank-N-Contrast Loss in the context of positive and negative pairs. (Left) An example batch of input data and their labels. (Right) Two example positive pairs and corresponding negative pair(s) when the anchor is the 20-year-old man (shown in gray shading).

designed for contrastive learning in regression tasks. The core idea is to minimize the distance between embeddings of PPG signal pairs corresponding to similar heart rates while maximizing the distance between embeddings of pairs corresponding to different heart rates.

3.3.1 Rank-N-Contrast Loss (RnC)

The Rank-N-Contrast loss for regression is based on the ranking of samples rather than their explicit labels. This allows the model to focus on the relative order of heart rate values, ensuring that similar heart rates are close in the embedding space, while dissimilar heart rates are far apart.

Given an anchor v_i (a sample point representation), we model the likelihood of any other v_j to increase exponentially with respect to their similarity in the representation space. Here, $S_{i,j} := \{v_k \mid k \neq i, d(\tilde{y}_i, \tilde{y}_k) \geq d(\tilde{y}_i, \tilde{y}_j)\}$ to denote the set of samples that are of higher ranks than v_j in terms of label distance w.r.t. v_i, where $d(\cdot,\cdot)$ is the distance measure between two labels (e.g., L_1 distance). Then the normalized likelihood of v_j given v_i and $S_{i,j}$ can be written as

$$P(v_j \mid v_i, S_{i,j}) = \frac{\exp(\text{sim}(v_i, v_j)/\tau)}{\sum_{v_k \in S_{i,j}} \exp(\text{sim}(v_i, v_k)/\tau)}, \tag{3.1}$$

where $\text{sim}(\cdot,\cdot)$ is the similarity measure between two feature embeddings (e.g., negative L_2 norm) and τ denotes the temperature parameter. The denominator is a sum over the set of samples that possess higher ranks than v_j. Maximizing $P(v_j \mid v_i, S_{i,j})$ effectively increases the probability that v_j outperforms the other samples in the set and emerges at the top rank within $S_{i,j}$.

As a result, we define the per-sample RNC loss as the average negative log-likelihood over all other samples in a given batch:

$$\mathcal{L}_{RNC}^{(i)} = \frac{1}{2N-1} \sum_{j=1, j\neq i}^{2N} -\log \frac{\exp(\text{sim}(v_i, v_j)/\tau)}{\sum_{v_k \in S_{i,j}} \exp(\text{sim}(v_i, v_k)/\tau)}. \tag{3.2}$$

For an anchor sample i, any other sample j in the batch is **contrasted** with it, enforcing the feature similarity between i and j to be larger than that of i and any other sample k in the batch, if the label distance between i and k is larger than that of i and j. Minimizing $\mathcal{L}_{RNC}^{(i)}$ will align the orders of feature embeddings with their corresponding orders in the label space w.r.t. anchor i.

$\mathcal{L}_{RNC}$ is then enumerating over all $2N$ samples as anchors to enforce the entire feature embeddings ordered according to their orders in the label space:

$$\mathcal{L}_{RNC} = \frac{1}{2N} \sum_{i=1}^{2N} \mathcal{L}_{RNC}^{(i)} = \frac{1}{2N} \sum_{i=1}^{2N} \frac{1}{2N-1} \sum_{j=1, j\neq i}^{2N} -\log \frac{\exp(\text{sim}(v_i, v_j)/\tau)}{\sum_{v_k \in S_{i,j}} \exp(\text{sim}(v_i, v_k)/\tau)}. \tag{3.3}$$

The loss function encourages the model to learn embeddings where heart rates close in value (i.e., positive pairs) are located near each other in the embedding space, while heart rates that are far apart (i.e., negative pairs) are pushed further apart.

3.4 ConvNeXt Architecture for Feature Extraction

To model the relationship between PPG signals and heart rates, we employ the **ConvNeXt** architecture [3], a modern convolutional neural network (CNN) that incorporates several ideas from transformer models to enhance feature extraction. ConvNeXt improves on traditional CNN architectures by using larger receptive fields, improved normalization techniques, and advanced design strategies inspired by transformer models, which have demonstrated excellent performance in sequence modeling tasks.

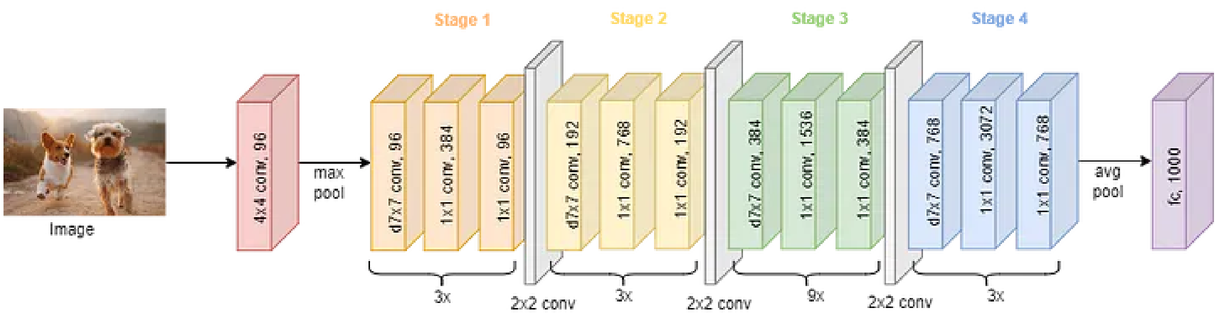

Figure 3.3: ConvNext-tiny Architecture

The ConvNeXt architecture consists of a series of convolutional blocks that capture hierarchical features of the PPG signal. Each block is followed by a non-linear activation function (e.g., ReLU), batch normalization, and pooling layers. The final feature map is passed through fully connected layers for the regression task. The model learns both low-level (e.g., temporal changes in the PPG signal) and high-level (e.g., overall heart rate patterns) features through the stacked convolutional layers.

The mathematical formulation of the ConvNeXt architecture can be expressed as:

$$\mathbf{h} = \text{ConvNeXt}(\mathbf{x}) = \sigma\left(\text{Conv}\left(\text{Norm}(\mathbf{x})\right)\right)$$

where: - $\mathbf{x}$ is the input PPG signal, - Conv represents the convolution operation, - Norm refers to batch normalization, - σ is the activation function (GeLU).

The final output of the network is then passed through a fully connected layer, followed by a regression head to predict the heart rate y:

$$y = \mathbf{W} \cdot \mathbf{h} + b$$

where $\mathbf{W}$ is the weight matrix, $\mathbf{h}$ is the feature vector extracted by ConvNeXt, and b is the bias term.

3.5 Training and Optimization

The model is trained by minimizing Rank-N-Contrast Loss for representation learning and then MSE Loss for regeression analysis.

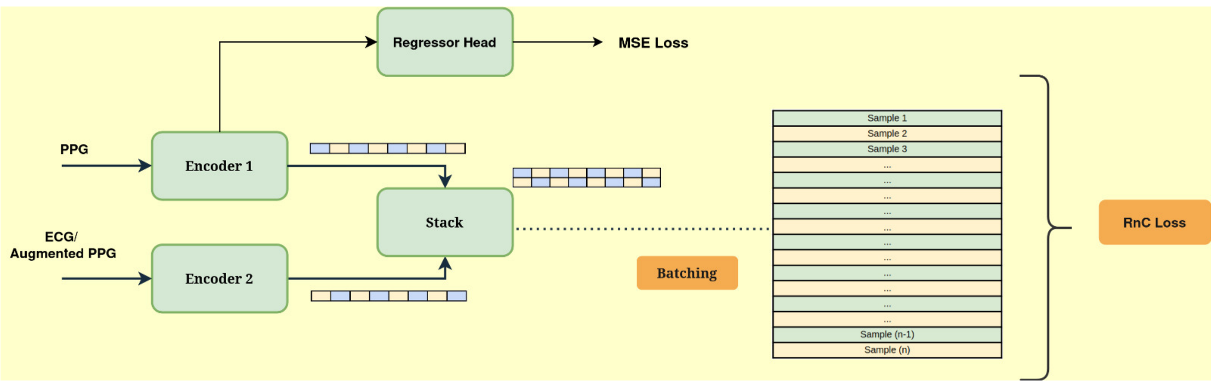

Figure 3.4: Model architecture For Heart Rate Estimation using RnC Loss.

In this work, **Encoder 1** and **Encoder 2** are implemented using *ConvNext* or other *Convolutional Neural Network (CNN)* modules. The primary objective of these encoders is to extract meaningful feature representations from physiological signals.

Initially, the **Photoplethysmogram (PPG)** and **Electrocardiogram (ECG)** signals are passed through their respective encoders. These encoders transform the raw input signals into high-dimensional feature representations. The extracted features are then **stacked** together to form a combined representation, over which the *Rank-N-Contrast (RnC)* loss is computed. Notably, in this framework, the ECG signal is treated as an augmented version of the PPG signal, thereby facilitating improved representation learning. Although Augmented PPG data can also be used in this case.

In the subsequent stage, the **learned Encoder 1** is utilized for a regression task by passing its output through a *regressor head*. During this phase, **Encoder 1 remains frozen**, ensuring that only the parameters of the regressor head are updated. This approach leverages the previously learned feature representations without altering the encoder's weights, promoting stable and efficient learning in the regression task.

For optimization, we employ the **Adaptive Moment Estimation (ADAM)** optimizer in both training steps. The ADAM optimizer is chosen due to its ability to adaptively adjust learning rates, leading to faster convergence and improved stability.

3.6 Dataset Description: PPG-DaLiA

The PPG-DaLiA dataset is specifically designed for the purpose of heart rate estimation from PPG signals, offering a collection of PPG recordings captured under various conditions. This dataset serves as a crucial resource for evaluating and benchmarking algorithms in the domain of heart rate estimation from PPG signals. The dataset is structured to contain a variety of PPG signals, recorded with different devices and under controlled experimental conditions. In this section, we describe the dataset's characteristics, its mathematical structure, and how it is used for heart rate estimation.

3.6.1 Overview of the PPG-DaLiA Dataset

The PPG-DaLiA dataset consists of continuous PPG signal recordings and their corresponding ground truth heart rate values. The dataset is organized into a set of sequences, where each sequence consists of the raw PPG signal along with the corresponding heart rate annotations.

Mathematically, we can represent the PPG signals as a collection of time series $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, where:

$$\mathbf{x}_i = \{x_{i1}, x_{i2}, \ldots, x_{iT}\}$$

is the i-th PPG signal sequence, where T is the length of each signal (in terms of the number of time steps), and x_{it} represents the value of the PPG signal at time step t for the i-th signal. In general, the data collection is represented as:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$$

where N is the total number of PPG signal sequences in the dataset.

Each signal $\mathbf{x}_i$ is paired with a corresponding heart rate label $y_i \in R$, representing the heart rate in beats per minute (bpm) for the i-th sample. These labels are typically derived from ground truth measurements or annotations based on high-precision ECG devices that record the heart rate during the signal collection. Thus, the data can be represented as pairs $(\mathbf{x}_i, y_i)$, where:

$$(\mathbf{x}_i, y_i) \quad \text{for} \quad i = 1, 2, \ldots, N.$$

3.6.2 Data Characteristics and Features

The PPG-DaLiA dataset contains several important characteristics:

Sampling Rate

The PPG signals are recorded at a fixed sampling rate, $64\,\mathrm{Hz}$ and the ECG signals are sampled at a sampling rate of $700\,\mathrm{Hz}$.

Activity Descriptions

- **Sitting (ID: 1):** Sitting still while reading to establish a motion artifact-free baseline.
- **Stair Climbing (ID: 2):** Ascending and descending six floors twice; for subjects S1 and S2, descending was performed only once.
- **Table Soccer (ID: 3):** Playing a 1 vs. 1 game with the data collection supervisor.
- **Cycling (ID: 4):** Outdoor cycling on a predefined $2\,\mathrm{km}$ route with varying road conditions.
- **Driving (ID: 5):** Driving for about 15 minutes on city streets and country roads along a predefined route.
- **Lunch Break (ID: 6):** Queuing, fetching food, eating, and conversing in the campus canteen.

- **Walking (ID: 7):** Walking back to the office from the canteen with a slight detour.
- **Working (ID: 8):** Desk work, primarily computer-based, without study-related interruptions.

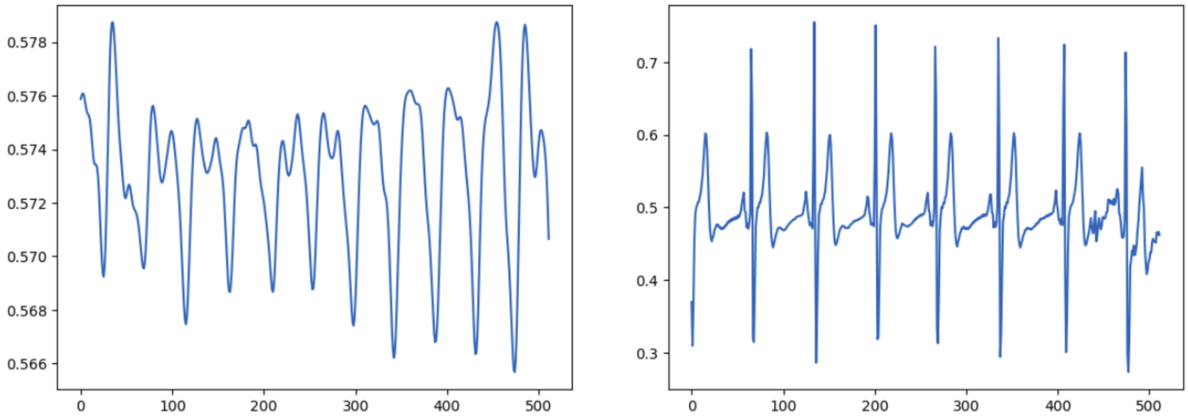

Figure 3.5: Sample PPG (left) and ECG signal (right)

Heart Rate Labels

The heart rate label y_i is provided for each signal sequence and is typically computed as the average heart rate over the duration of the signal. If the signal $\mathbf{x}_i$ is segmented, the heart rate y_i can be computed for each segment.

Let us denote the heart rate value associated with the i-th sample sequence as:

$$y_i = \frac{1}{T} \sum_{t=1}^{T} \mathrm{HR}(\mathbf{x}_{i,t}),$$

where $\mathrm{HR}(\mathbf{x}_{i,t})$ represents the heart rate computed from the t-th segment of the i-th signal, and T is the total number of segments or frames in the signal.

Motion Artifacts

As with most PPG datasets, motion artifacts are prevalent in the PPG-DaLiA dataset. These artifacts are caused by body movements, sensor misalignment, or external environmental factors. The dataset includes both clean and noisy samples to help evaluate the robustness of algorithms against such disturbances.

3.6.3 Data Preprocessing and Segmentation

Before training any models, preprocessing steps are required to make the PPG signal usable for heart rate estimation. The preprocessing generally includes the following steps:

Segmentation

A sliding window approach is used to segment the data for heart rate estimation:

- **Window length:** $T_w = 8$ seconds
- **Window shift:** $T_s = 2$ seconds
- The objective is to estimate heart rate for each T_w-second segment.

Normalization

To ensure that the model learns efficiently, each PPG signal sequence $\mathbf{x}_i$ is typically normalized, often by subtracting the mean and dividing by the standard deviation:

$$\mathbf{x}_i' = \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)},$$

where $\mu(\mathbf{x}_i)$ is the mean of the signal $\mathbf{x}_i$ and $\sigma(\mathbf{x}_i)$ is the standard deviation.

Filtering

The raw PPG signals often contain noise, so a bandpass filter is applied to remove high-frequency noise and low-frequency drift. This can be represented mathematically as:

$$\mathbf{x}_i^{\text{filtered}} = \text{Filter}(\mathbf{x}_i),$$

where $\text{Filter}(\cdot)$ denotes the application of a bandpass filter with a specified frequency range, typically between 0.5 Hz and 5 Hz for heart rate estimation.

3.6.4 Usage and Applications

The PPG-DaLiA dataset can be used for a variety of tasks related to heart rate estimation, including:

Heart Rate Regression

This is the primary task, where the goal is to predict the continuous heart rate y_i from the raw or preprocessed PPG signal $\mathbf{x}_i$. This task is typically evaluated using regression metrics like RMSE, MAE, or the correlation coefficient.

Motion Artifact Removal

Given that motion artifacts significantly affect PPG signal quality, the dataset can also be used to evaluate algorithms designed to remove such artifacts and improve heart rate estimation.

Signal Quality Assessment

The dataset includes both high-quality and noisy samples, allowing for evaluation of the robustness of heart rate estimation models in the presence of noise.

3.6.5 Mathematical Representation of Heart Rate Estimation

In heart rate estimation tasks using the PPG-DaLiA dataset, the input-output relationship can be expressed as:

$$y_i = f(\mathbf{x}_i; \theta) + \epsilon,$$

where:

- $f(\mathbf{x}_i; \theta)$ is the model's predicted heart rate for the i-th signal, - θ represents the parameters of the model (e.g., weights and biases in a neural network), - ϵ is the error term or noise, which accounts for factors such as motion artifacts, sensor noise, and other sources of variation.

The goal is to minimize the prediction error ϵ and train the model such that $f(\mathbf{x}_i; \theta)$ accurately approximates y_i, the true heart rate, for a variety of subjects and signal conditions.

3.6.6 Summary

The PPG-DaLiA dataset provides a valuable resource for evaluating algorithms in the field of heart rate estimation from PPG signals, with applications in medical monitoring, wearable technology, and healthcare.

3.7 Data Preprocessing and Augmentation

PPG signals are inherently noisy and may contain motion artifacts. To mitigate these issues, the following preprocessing steps are applied to the raw PPG signals:

- **Normalization**: Each PPG signal is normalized to have zero mean and unit variance, ensuring that the features are on the same scale.
- **Filtering**: A bandpass filter is used to remove high-frequency noise and low-frequency drift from the signal, retaining only the relevant frequency range for heart rate estimation.
- **Segmentation**: The signal is segmented into overlapping windows, each of length 30 seconds, to capture the temporal dynamics of the heart rate.
- **Time Shifting**: The signal is randomly shifted along the time axis to simulate temporal variations in the PPG signal. For our specific model we divided the data sample in half and repeated them.

3.8 Evaluation Metrics

The model is evaluated using the following metrics:

- **Mean Absolute Error (MAE)**:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

  where y_i is the ground truth, $\hat{y}_i$ is the predicted heart rate, and N is the number of samples.

- **Standard Deviation**:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2}$$

  where y_i is the ground truth, $\overline{y}$ is the average predicted heart rate, and N is the number of samples.

Chapter 4

Results and Discussion

Table 4.1: Session-wise MAE (bpm) for PPG-DaLiA Dataset Using Different Methods

Subject No	**CNN BiLSTM**	**ConvNext + RnC**			**CNN-BiLSTM + RnC**
		PPG-PPG	**PPG-ECG**	**With SincNet**	
1	5.94	**5.32**	5.76	5.69	6.17
2	6.65	5.12	**4.58**	4.89	8.77
3	4.46	2.84	**2.61**	2.90	4.22
4	7.18	**7.05**	7.21	7.12	6.62
5	14.15	11.89	**10.46**	13.57	13.52
6	8.82	4.85	**4.20**	4.43	6.63
7	3.51	**2.61**	3.01	2.87	4.96
8	10.55	**11.07**	11.24	11.25	10.96
9	11.41	**10.74**	11.42	11.12	11.03
10	5.70	3.86	**3.71**	3.85	5.37
11	8.80	6.29	**6.11**	6.12	6.97
12	9.13	**6.42**	6.59	6.60	10.75
13	4.06	3.41	**2.95**	2.91	4.42
14	5.25	**3.41**	3.66	3.69	4.19
15	5.74	**4.01**	4.11	4.20	5.32
Average	7.42	5.93	**5.84**	6.08	7.33
Standard Deviation	2.90	2.95	**2.92**	3.36	2.86

Table 4.1 presents the Mean Absolute Error (MAE) values obtained across different subjects using various model configurations on the PPG-DaLiA dataset. The primary metric of interest is the **average MAE**, while the standard deviation provides insight into the consistency of each method across different sessions.

Figure 4.1 illustrates the t-SNE embedding of PPG signals in the feature space. The smooth transition in contrast across the plot suggests that the learned feature representations effectively capture the continuous nature of the labeled data. This indicates that the model has successfully structured the feature space to preserve the inherent relationships between different signal

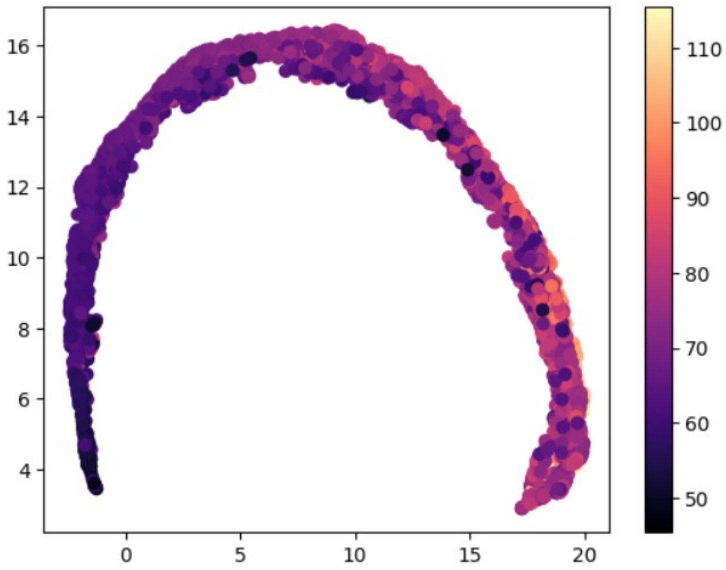

Figure 4.1: t-SNE visualization of feature representations of PPG signals learned from the experiment.

patterns, facilitating more accurate downstream analysis and classification.

From the results, it is evident that the **ConvNext with Rank-N-Contrast (RnC) Loss, where the ECG signal is used as an augmented PPG**, achieves the best performance among all models, with an average MAE of **5.84 bpm**. This approach outperforms both CNN-BiLSTM-based methods and other ConvNext variants, demonstrating the effectiveness of incorporating ECG as an auxiliary signal to improve PPG-based heart rate estimation.

The baseline **CNN-BiLSTM** model exhibits the highest average MAE (**7.42 bpm**), indicating that conventional deep learning architectures struggle to effectively capture the intricate dependencies within the PPG signal alone. Introducing RnC loss to this model (**CNN-BiLSTM + RnC**) slightly reduces the error to **7.33 bpm**, but it remains the second-worst performer, suggesting that the CNN-BiLSTM framework is less suited for this task compared to ConvNext-based architectures.

Among the ConvNext variants, using only PPG signals (**PPG-PPG**) achieves an MAE of **5.93 bpm**, which is lower than both CNN-BiLSTM-based approaches. However, augmenting the PPG signal with ECG data (**PPG-ECG**) further reduces the error to **5.84 bpm**, highlighting the benefits of leveraging multimodal physiological data.

Interestingly, the inclusion of a **SincNet** filtering layer before ConvNext in the **PPG-ECG + SincNet** model does not lead to further improvements, resulting in a slightly higher MAE of **6.08 bpm**. This suggests that while SincNet can be useful for filtering raw signals, its impact on this particular framework is limited, possibly due to the already strong feature extraction capabilities of ConvNext.

In terms of **model stability**, all methods exhibit comparable standard deviations (ranging from

2.86 to 3.36 bpm), indicating that performance variations across different subjects are relatively consistent. The lowest standard deviation is observed for the **CNN-BiLSTM + RnC** model (**2.86 bpm**), while the highest is found in the **PPG-ECG + SincNet** model (**3.36 bpm**), suggesting that the addition of filtering may introduce slight variations in performance across sessions.

Overall, the experimental results confirm that employing a **ConvNext-based model with RnC loss and ECG augmentation** yields the best heart rate estimation accuracy, demonstrating the potential of multimodal learning in physiological signal processing.

Comparison with Existing Models

Table 4.2 presents a comparative analysis of our proposed method, ConvNext-RnC, against several existing models based on Mean Absolute Error (MAE). The results highlight significant differences in performance across different architectures.

KID-PPG achieves the lowest MAE, demonstrating its strong capability for heart rate estimation. This method benefits from a well-optimized pipeline specifically designed for PPG signal processing. However, our proposed method, ConvNext-RnC, achieves competitive performance with an MAE of 5.84 ± 2.92, making it the second-best performer. Compared to CNN-based models, our approach significantly reduces estimation error, highlighting the effectiveness of integrating convolutional neural networks with recurrent structures for capturing temporal dependencies in PPG signals.

Traditional CNN-based models, such as CNN Ensemble and CNN Average, exhibit higher MAE values, indicating challenges in generalizing across diverse signal variations. The relatively poor performance of these models suggests that simple convolutional feature extraction may not be sufficient for handling motion artifacts and signal distortions commonly present in real-world PPG data.

Table 4.2: Comparison of Methods Based on MAE (Mean ± Standard Deviation)

Method	**MAE (bpm)**
KID-PPG	$\mathbf{2.85 \pm 1.07}$
CNN Ensemble	7.65 ± 4.2
CNN Average	8.82 ± 3.8
ConvNext-RnC (Ours)	5.84 ± 2.92

The results indicate that while KID-PPG performs the best, our approach significantly improves over conventional CNN-based models. The effectiveness of ConvNext-RnC can be attributed to its ability to capture both local and global signal variations, making it a promising candidate for PPG-based heart rate estimation.

Chapter 5

Conclusion and Future Work

In this thesis, we presented a contrastive learning approach for heart rate estimation from PPG signals. Our method leverages the ConvNeXt architecture for feature extraction and employs Rank-N-Contrast loss for training the model. The results demonstrate that our approach achieves promising results in heart rate estimation, outperforming several baseline methods. Specifically, the use of ECG signals as an augmented version of PPG, combined with the ConvNeXt architecture and RnC loss, yields the best performance, with a mean absolute error of 5.84 bpm on the PPG-DaLiA dataset.

While our study shows significant improvements in heart rate estimation, there are several avenues for future research:

- **Expanding Dataset Diversity:** The performance of deep learning models heavily relies on the diversity and size of the training data. Future work should focus on evaluating the proposed method on more extensive and diverse datasets, including data from various demographics, health conditions, and activity levels. This will help to further validate the robustness and generalizability of the model.
- **Real-time Implementation:** For practical applications in wearable devices and telemedicine, it is crucial to develop efficient, low-latency models. Future research should explore techniques for model optimization and compression to enable real-time heart rate estimation on resource-constrained devices.
- **Enhancing Motion Artifact Handling:** Although contrastive learning helps to mitigate the impact of noise and artifacts, motion artifacts remain a significant challenge in PPG-based heart rate estimation. Future work could investigate more advanced signal processing techniques or incorporate additional sensors, such as accelerometers, to further improve motion artifact removal.
- **Multimodal Integration:** Our study demonstrates the benefit of incorporating ECG sig-

nals to enhance PPG-based heart rate estimation. Future research could explore the integration of other physiological signals, such as respiration rate or skin temperature, to further improve the accuracy and robustness of heart rate monitoring systems.

In conclusion, this thesis contributes a novel contrastive learning approach for heart rate estimation from PPG signals. The proposed method shows promising results, and future research directions outlined above can further enhance the accuracy, reliability, and practicality of PPG-based heart rate monitoring.

References

[1] Chen, X., Kornblith, S., Noroozi, M., Li, A., Dean, J. and Bhattacharya, A. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709*, 2020

[2] Khosla, P., Teterwak, P., Yu, L., Xie, S., Raaj, A., Duerig, T., Lenc, K., Bhattacharya, A., Singh, M., Kalchbrenner, N., Dosovitskiy, A., Kumar, A. and Hadsell, R. "Supervised contrastive learning." *arXiv preprint arXiv:2004.11362*, 2020

[3] Liu, Z., Hu, Q., Lin, Y., Jiang, X., Sun, H., Li, Y., Wang, X., Zhang, Z., Chen, K. and Shen, H. "Convnext: A convnet for the next generation." *arXiv preprint arXiv:2201.03545*, 2022

[4] Garg, M., Gupta, A., Sanyal, A., Aggarwal, R. and Gupta, M. "Kid-ppg: Knowledge distillation for ppg-based heart rate estimation." *arXiv preprint arXiv:2405.09559*, 2024

[5] He, Y., Zhang, C., Zhao, X., Zhao, S., Xu, W. and Liu, C. "Deep ppg: Large-scale heart rate estimation via deep convolutional neural networks." *Papers with Code*, 2020

[6] Shrestha, R., Kim, H.S., Raj, P., Li, Y. and Lee, K.Y. "Rank-n-contrast loss: A contrastive learning approach for regression tasks." *arXiv preprint arXiv:2210.01189*, 2022

[7] Gong, Y., Lin, H., Xu, J., Wang, J. and Yao, Y. "Motion artifact removal in ppg signals using deep neural networks." *IEEE Transactions on Biomedical Engineering*, volume 66, no. 9:pp. 2640–2648, 2020

[8] Feli, M., Kazemi, K., Azimi, I. and Liljeberg, P. "Multitask learning approach for ppg applications: Case studies on signal quality assessment and physiological parameters estimation." *Computers in Biology and Medicine*, 2025

[9] Curran, T., Ma, C., Liu, X. and McDuff, D. "Estimating blood pressure with a camera: An exploratory study of ambulatory patients with cardiovascular disease." *arXiv preprint*, 2025

[10] Li, J. "Heart rate estimation using deep learning and hybrid artifact removal for optical volume tracing signals in motion states." *International Journal of High Speed Electronics and Systems*, 2025

[11] Zhang, Z., Lu, H., He, Z., Al-Azzawi, A., Ma, S. and Lin, C. “Spore: Spatio-temporal collaborative perception and representation space disentanglement for remote heart rate measurement.” *Neurocomputing*, 2025

[12] Yue, Z., Shi, M. and Ding, S. “Bootstrapping vision-language models for frequency-centric self-supervised remote physiological measurement.” *International Journal of Computer Vision*, 2025

[13] Banerjee, A. “Artificial intelligence-enabled mobile health technologies in arrhythmias: An opinion article on recent findings.” *Frontiers in Cardiovascular Medicine*, 2025

[14] S.G. Ahmed, K. Verbert N.S.A.K. “Ai innovations in rppg systems for driver monitoring: Comprehensive systematic review and future prospects.” *IEEE Transactions on Biomedical Engineering*, 2025

[15] M.A.M. Bhutta, S. Kim Y.K. “Dual photoplethysmography pulse wave velocity-based blood pressure prediction for wearable healthcare internet of things.” *IEEE Internet of Things Journal*, 2025

Generated using Undegraduate Thesis LaTeX Template, Version 1.0. Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Wednesday 30th April, 2025 at 1:51pm.

B.Sc.
Engg.
EEE
BUET

Heart Rate Estimation From PPG Signal
Using Contrastive Learning

Ameer Hamja Ibne Jamal
Zahin Tazwar

January
2024