

Literature Review for Year 3 Group Project

Harsh Agrawal¹

¹*Department of Bioengineering, Imperial College London*

November 24, 2024

1. A Self-Supervised Algorithm for Denoising Photoplethysmography Signals for Heart Rate Estimation from Wearables

Pranay Jain, Cheng Ding, Cynthia Rudin, Xiao Hu

In their paper, Pranay et al. introduce SPEAR (Self-supervised PPG Erase Artifacts and Reconstruct), a novel contrastive learning framework designed to denoise photoplethysmography (PPG) signals for heart rate estimation from wearable devices. PPG signals are frequently compromised by noise from motion artifacts and factors like sweat, pressure, or irregular contact with the sensor. Traditional denoising methods often underperform, either failing to remove all noise or inadvertently discarding valuable information from uncorrupted portions of the signal, leading to reduced accuracy in downstream applications.

SPEAR addresses these challenges using a denoising autoencoder trained with a contrastive loss function. The encoder consists of four convolutional layers, each followed by batch normalization and ReLU activation. The model is trained on 30-second PPG signals sampled at 64 Hz, where artificial noise is introduced by partially corrupting the input signals. The autoencoder reconstructs clean signals, with the loss calculated as the root mean squared error (RMSE) between the original and reconstructed signals.

Trained on a dataset of approximately 62,000 PPG signals, SPEAR demonstrates state-of-the-art performance in heart rate estimation on the test set. The authors argue that SPEAR can serve as an effective preprocessing step for any application relying on PPG signals, enhancing signal quality and improving downstream analyses. Moreover, the use of large-scale self-supervised pre-training eliminates the need for extensive labeled datasets, enabling SPEAR to be applied to most heterogeneous unlabeled datasets. It also holds significant potential for extending its utility to electrocardiography (ECG) signal analysis which is still unexplored.

2. State-of-the-Art of Stress Prediction from Heart Rate Variability Using Artificial Intelligence

Yeaminul Haque, Rahat Shahriar Zawad, Chowdhury Saleh Ahmed Rony, Hasan Al Banna, Tapotosh Ghosh, M. Shamim Kaiser, and Mufti Mahmud

The authors of this meta-review systematically examine 43 studies that apply various AI algorithms to heart rate variability (HRV) data for stress prediction. They categorize these methodologies into machine learning (ML) and deep learning (DL) approaches, outlining the strengths and limitations of each. Traditional ML models, such as support vector machines (SVM) and random forests (RF), offer interpretability and computational efficiency but often struggle to handle the non-linear and complex patterns present in HRV data. Conversely, DL architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), excel at capturing intricate temporal features of HRV signals, though they come with increased computational complexity and reduced interpretability.

The review also highlights the critical importance of preprocessing techniques, such as noise reduction, feature extraction, and normalization, in enhancing model accuracy. Time-domain, frequency-domain, and non-linear features are frequently extracted to serve as inputs for AI models. Additionally, the authors address challenges in HRV-based stress prediction, including individual variability, the influence of confounding factors like physical activity and circadian rhythms, and the lack of large, diverse datasets for robust model training and validation.

A concerning highlight of this review was that most ML models report unrealistically high accuracy (95%) on HRV benchmarks. This suggests potential overfitting due to small training datasets or inadequacies in benchmark rigor. The authors call for improved approaches, such as self-supervised pre-training, to address overfitting and enhance model generalizability. This could pave the way for more reliable and robust HRV-based stress prediction systems, making it a promising direction for future research.

3. Real-time machine learning model to predict in-hospital cardiac arrest using heart rate variability in ICU

Hyeonhoon Lee, Hyun-Lim Yang, Ho Geol Ryu, Chul-Woo Jung, Youn Joung Cho, Soo Bin Yoon, Hyun-Kyu Yoon, and Hyung-Chul Lee

In their study, the authors present a real-time machine learning model designed to predict in-hospital cardiac arrest (IHCA) by analyzing heart rate variability (HRV) in intensive care unit (ICU) patients. Recognizing that traditional HRV metrics—such as the standard deviation of normal RR intervals (SDNN) and low-frequency (LF) and high-frequency (HF) powers—hold predictive value for IHCA, they note that existing methods often focus on isolated factors, thereby missing the comprehensive insights offered by a broader range of HRV measures.

To address this, the researchers developed a Light Gradient Boosting Machine (LGBM) model incorporating 33 distinct HRV features. They trained this model on a dataset comprising 3,861 patients, employing hyperparameter optimization to determine the optimal configuration, followed by five-fold cross-validation to ensure robustness. The model achieved an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.88, indicating strong predictive performance.

Post-training analysis of feature importance, utilizing SHAP (Shapley additive explanations) values, revealed that the most influential predictor was the baseline width of the triangular interpolation of the RR interval histogram (TINN). This was followed by the HRV triangular index (HTI) and the inverse of the average length of acceleration and deceleration segments (IALS).

Notable limitations of this study were its exclusion of treatment interventions' effects on patient outcomes while focusing solely on baseline predictors and poor performance when in case forecasting events occurred within a 24-hour window. This omission could lead to potential false positives, reinstating the necessity for more comprehensive metrics and controls to validate the efficacy of HRV-based predictive models.

4. Learning Transferable Visual Models From Natural Language Supervision

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever

In contrast to other studies in the literature review, this review is not related to HRV but solely focuses on an ML framework namely CLIP (Contrastive Language-Image Pretraining), a self-supervised learning framework that our group considers highly applicable for large-scale pretraining of HRV-based models.

Released by OpenAI in 2019, CLIP is designed to learn transferable (or joint) embeddings for images and text by leveraging a vast dataset of 400 million image-caption pairs. The architecture comprises two encoders: a 63M-parameter 12- layer transformer-based text encoder and a set of vision transformer-based image encoder (ViT-B/32, a ViT-B/16, and a ViT-L/14). During training, the model processes batches of image-caption pairs, encoding them separately through the respective encoders to generate embeddings. A contrastive loss function is employed to minimize the vector distance between embeddings of matching image-caption pairs while maximizing the distance between mismatched pairs. Practically, this is achieved using a cross-entropy loss over the dot product of the embeddings.

Even with such a simple training paradigm CLIP is able to significantly outperform state-of-the-art models on various ImageNet benchmarks, including object recognition tasks, while being more efficient in terms of speed and size. This is primarily because CLIP is able to efficiently leverage the large weakly-labelled pool of dataset for pre-training. CLIP is one of the strongest demonstrations of the effectiveness of self-supervised contrastive learning models among all domains. CLIP, and other contrastive learning methods still remain under-explored in the domain of HRV, where there is a pronounced scarcity of large labeled datasets, highlighting its potential as a signal reconstruction or a generic pre-training framework in our project.