

Modelling Naturalistic Work Stress Using Spectral HRV Representations and Deep Learning

Juan Lorenzo Hagad^{*1} Ken-ichi Fukui^{*2} Masayuki Numao^{*2}

^{*1}Graduate School of Information Science and Technology

^{*2} Institute of Scientific and Industrial Research
Osaka University, Japan

With the proliferation of wearable devices and the inflow of new health data, artificial intelligence is expected to revolutionize the field of wellness and health management by providing potential tools for analyzing harmful conditions like prolonged stress. Currently, one of the standard measurements used by medical practitioners to measure stress is heart rate variability (HRV), a set of numerical indices that reflect autonomic balance. However, recent advances in machine learning have shown that learned features tend to outperform hand-crafted features. In this work we propose a more expressive intermediate data representation based on Lomb-Scargle periodograms combined with the feature learning capabilities of deep learning. Using stress data from naturalistic work activities, we tested different shallow and deep learning architectures and show that significant improvements can be achieved compared to traditional HRV indices. Results show that models trained on our spectral-temporal representation significantly outperform models trained on traditional HRV indices for predicting naturalistic work stress.

1. Introduction

Findings from a large number of clinical studies [5, 8] have linked cardio-vascular disorders to psychological stressors. In particular, long-term exposure to stressful environments, such as those commonly found in offices, has been shown to be a major risk factor for major depressive disorder (MDD), a disorder that affects an estimated 350 million people across the world [9]. In response to this, we have witnessed a surge in the development of ambient intelligences for the purpose of mental health and stress monitoring [1].

1.1 HRV Stress Analysis

Historically, heart rate variability (HRV) spectrograms were used by medical practitioners as the primary tool to measure cardiac neural regulation externally. This was brought out by findings that mental and physical states have certain effects on cardiac rhythm that could easily and unintrusively be measured with electrocardiographs (ECG) [6]. In the past few decades, studies have established the connection between HRV and the sympatho-vagal stress response as being visible in both the time and frequency domains. Through experiments and empirical analysis, these works have identified indices which show high correlations with physical and mental stress, particularly with the high and low frequency ranges [10]. However, even now there is no widely accepted standard for stress evaluation based on HRV features. In recent surveys [14, 7] it has been found that neurobiological evidence links some components of HRV to activity in cortical regions related to stress appraisal as well as other brain mechanisms. As it stands, it seems that one of the main challenges is identifying which components of HRV may be relevant to evaluating stress.

1.2 Deep Learning and Heart Rate Analysis

Traditionally, affective analysis such as those done for stress have mostly focused towards using visual or auditory measures [15], where deep learning has been found to be effective. In recent years, others works in the field of ECG analysis for medical applications have also pushed for end-to-end applications of deep learning. In one of the most prominent works [12] they used a 34-layer residual convolutional neural network (CNN) to analyze raw ECG sequences in order to identify 14 different types of arrhythmia. Using a very large dataset of over 60,000 recordings they were able to train a model that performed better on average than actual cardiologists recruited for the experiment. Their work demonstrates how deep learning is able to identify structural patterns in heart rate data. In our work, we attempt to exclude morphological patterns and instead focus on frequency-related features of heart beats. Towards this end we introduce the use of short-time Lomb-Scargle (LS) spectrograms to train CNNs.

2. Methodology

This work aims to use spectral HRV data representations together with deep learning models to predict mental stress. Specifically, we use a sliding window Lomb-Scargle approach to build a spectrogram of heart rate in lieu of the traditional short-term Fourier transform (STFT). Next, we use pre-built deep models designed for image recognition and compared them with more simplified architectures and train them to recognize stress. In the following sections we describe the experiments and tools used in this research.

2.1 Dataset

For the following experiments we use the same work stress dataset introduced in a previous work [4]. It consists of ECG recordings of subjects performing naturalistic desk work activities in front of a PC. The dataset was gathered

Contact: Juan Lorenzo Hagad, Department of Architecture for Intelligence, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan, +81-6-6879-8426, hagad@ai.sanken.osaka-u.ac.jp

from four subjects each contributing five separate hour-long sessions. This provided us with a total of 20 hours of ECG data that was used for training and testing.

2.2 Signal Processing

Previous works on spectral analysis of HRV applied the fast Fourier transform (FFT) [17] to produce periodograms from which the power of different frequency bands can be measured and compared. However, one limitation of the FFT algorithm is that it assumes that all samples used for analysis are evenly spaced in time. This poses a fundamental issue for heart rate since the RR intervals used to represent detected heart beats are not evenly sampled. A study [2] has shown that data resampling methods used to align RR interval data for FFT analysis can lead to artifacting and higher errors in the final PSD estimation. A better alternative has been put forward in the form of the Lomb-Scargle periodogram, a method based on FFT that does not require evenly sampled data.

2.2.1 Lomb-Scargle Periodogram

Periodograms are some of the most basic tools used for spectral analysis, the most common of which is the discrete Fourier transform (DFT). The DFT can be defined for any sampled dataset, $X(t_i), i = 1, 2, \dots, N_0$ as,

$$FT_X(\omega) = \sum_{j=1}^{N_0} X(t_j) \exp(-i\omega t_j). \quad (1)$$

The periodogram can then be defined as,

$$\begin{aligned} P_X(\omega) &= \frac{1}{N_0} |FT_X(\omega)|^2 \\ &= \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X(t_j) \exp(-i\omega t_j) \right|^2, \\ &= \frac{1}{N_0} \left[\left(\sum_j X_j \cos \omega t_j \right)^2 + \left(\sum_j X_j \sin \omega t_j \right)^2 \right] \end{aligned} \quad (2)$$

[3] which while sufficient for calculating an accurate estimation of the FFT of most signals was found to be sensitive to noisy signals and other artifacts [16]. In practice, it is necessary to apply additional spectral smoothing equations, such as *spectral window functions*, to reduce the variances of the spectrum. Most of these methods are designed to be applied on evenly sampled data. Though some may also be applied to the unevenly sampled case, such as in RR intervals, it has been found that it can lead to higher error and increased sensitivity to signal noise [11]. This can also be mitigated by the addition of more data for calculating the spectra as a way to taper off the signal-to-noise ratio, but this may not be an option for data-scarce or time-intensive domains. A reasonable alternative is the Lomb-Scargle periodogram [13], a notable method designed for use in calculating frequency spectrograms from unevenly sampled data streams such as heart beats. The equation is based on the FFT and is wholly similar but addresses some of the noise sensitivity and *spectral leakage* issues of the standard FFT

equation. Here, the periodogram is defined as,

$$P_X(\omega) = \frac{1}{2} \left\{ \frac{\left[\sum_j X_j \cos \omega(t_j - \tau) \right]^2}{\sum_j \cos^2 \omega(t_j - \tau)} \right\} + \frac{\left[\sum_j X_j \sin \omega(t_j - \tau) \right]^2}{\sum_j \sin^2 \omega(t_j - \tau)} \quad (3)$$

with τ defined as,

$$\tan(2\omega\tau) = \left(\sum_j \sin 2\omega t_j \right) / \left(\sum_j \cos 2\omega t_j \right). \quad (4)$$

This modified equation, outputs similar values to the original FFT in most cases, however it has a simpler statistical behavior and is equivalent to the reduction of the sum of squares in least-squares fitting of sin waves to data [13]. It is also time-translation invariant and reduces to a similar representation as DFT in cases where data is evenly spaced. This makes it ideal for HRV analysis and is the method used in this research. The next issue is how to tackle the non-stationary nature of heart rate. For this, we construct spectrograms from the periodograms.

2.2.2 Spectrogram Analysis

Spectrograms are visual representations of frequency spectra over time. They are tools that are used extensively in applications that deal with audio and radio signals. Recently, they have also seen frequent application in the analysis of physiological signals such as brainwaves and heart rate. The most common method used for building spectrograms is the short-time Fourier transform (STFT). Put simply it is built by applying the DFT over a pre-defined window of data and sliding the windowed function over the data series using a fixed amount of overlap. Formally, for any series $x(n)$ we can define the STFT as,

$$X_w(mS, \omega) = \sum_{n=-\infty}^{\infty} x(n)w(n - mS) \exp(-j\omega n) \quad (5)$$

for analysis window w , frame index m and step size S [18]. The internal equation $\sum_{n=-\infty}^{\infty} x(n)w(n - mS) \exp(-j\omega n)$ represents the Fourier transform of $x_w(mS, \omega)$ and is stored in the matrix \hat{X} whose columns are Fourier periodograms $P_X(\omega)$. To adapt the spectrogram to the unevenly sampled heart rate data, we replace this equation with the Lomb-Scargle equation,

$$\begin{aligned} X_w(mS, \omega) &= \frac{1}{2} \left\{ \frac{\left[\sum_j x(n) \cos \omega(t_j - \tau) \right]^2}{\sum_j \cos^2 \omega(t_j - \tau)} \right\} \\ &\quad + \frac{\left[\sum_j x(n) \sin \omega(t_j - \tau) \right]^2}{\sum_j \sin^2 \omega(t_j - \tau)} \end{aligned} \quad (6)$$

In this way we mitigate the sensitivity of the standard STFT while maintaining its ability to capture the nonstationary patterns in the temporal data. In effect, we end up with a very high dimensional representation that could potentially have more expressive and appropriate features for detecting stress. The next step is to use machine learning to discover these features and compare the performance

difference with models based on traditional HRV features. Due to the 2D spatial nature of data representation, convolutional neural networks should provide an efficient learning architecture.

3. Experiments and Validation

3.1 Dataset and Models

For the following experiments, we applied a sliding window method to the stress dataset in order to segment the data. This method also lends itself to real-time processing in the future. Through empirical testing, we found that 505-beat windows (6 mins) showed the best balance between stability and expressiveness while staying above the 5-minute minimum length recommended for short-term HRV analysis. We also applied a modal labelling scheme for segments which overlapped between multiple labels. However, we tweaked the unimodal threshold to 70% so segments with modal valued samples greater than the threshold could be labelled with the modal value, otherwise it would be marked as a "transitional" segment. The final experiment dataset was comprised of 1,500 windowed samples which were either labelled based on the 4-level stress annotation or as transitions.

For modeling, we trained perceptron and autoencoder models using the stress HRV data to serve as baselines and compared these to convolutional neural network (CNN) models trained on our spectral HRV data representation. The convolutional layers were topped with a 128-unit dense layer to match the ones used for the non-deep HRV stress models. For the CNN models, we tested both standard deep learning models typically used for image classification tasks as well as shallower CNN architectures to verify the hierarchical feature requirements of the dataset. Performance was validated using stratified 10-fold cross-validation.

3.2 Results and Discussion

First, we analyze the performance results for models trained on the standard HRV features, which were 14 time and frequency domain features, and compared them to the deep ResNet model trained on the LS spectrum data. As seen in Figure 1, the baseline HRV-MLP model achieved 52% test accuracy which was improved by up to 60% by applying denoising autoencoders for more efficient latent feature selection. These results are well-above the 20% random classification on this 5-class dataset.

3.2.1 Deep Spectral Model

Moving over to the spectral models, ResNet50 was able to achieve over 80% testing accuracy, which is well above both of the traditional HRV baselines. These significant gains could be explained by the effective combination of the more expressive spectral data with the feature learning ability of deep learning algorithms. The traditional indices used in traditional HRV analysis help humans understand physical conditions, but they may also obfuscate some intricate features that may be necessary for deeper affective analysis. By allowing the model to learn from a spatial-spectral representation, we are able to train a more effective model. However, this comes at the cost of increased com-

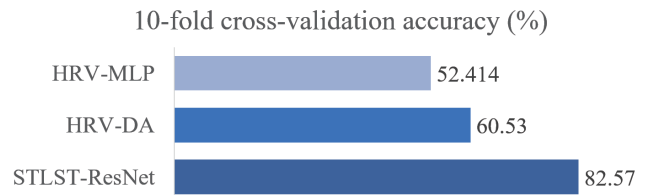


Figure 1: Comparison of test performance results between models trained on standard HRV features (MLP and DA) and the ResNet model trained on STLST heart rate spectrum data.

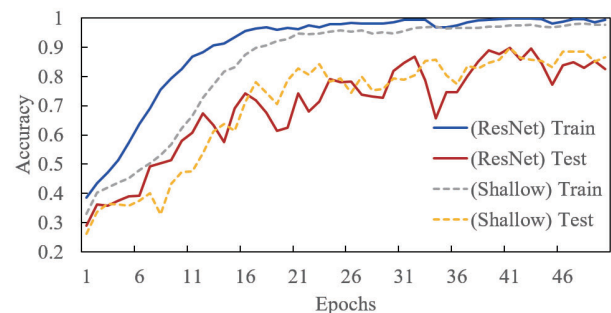


Figure 2: Learning and test performance of ResNet50 and the shallower 5-layer CNN model.

plexity since the standard 50-residual unit model requires over 25 million parameters. This not only leads to higher data requirements, but may also lead to more challenges in interpretation and optimization.

3.2.2 Shallow Spectral Model

To help alleviate the problem of increased complexity, we attempted to train a more compact model. For these tests, the goal was to reduce the depth and overall complexity of our model without compromising performance. This ultimately culminated in a model featuring 5 intermediate 128-filter CNN layers with BatchNorm for regularization. The relatively shallow model also meant residual units were unnecessary and training was much faster than the deeper model. It should also be noted that tests were also performed on shallower models below 5 layers, but this typically lead to larger performance variations and more noticeable overfitting. Figure 2 shows a comparison of the learning process of both the deep and shallow models. Despite using less than 1 million parameters, as opposed to the over 25 million of ResNet50, the shallow model is able to match the performance of the much larger ResNet model while also avoiding overfitting. Actual training time could also be reduced by up to 5x which may be useful in situations where occasional retraining is necessary.

4. Summary and Future Work

In this work we trained a deep model of work stress using a spectral representation of heart rate extracted using short-time Lomb-Scargle transform. The experiments showed that the deep models trained on the spectral rep-

resentation could easily outperform those trained on the traditional HRV features for detecting naturalistic stress levels. Additionally, we show that even relatively shallow models featuring at least 5 convolutional layers can match the performance of ResNet50, this could indicate that at least some shallow hierarchical features may be important for effectively modelling naturalistic stress from heart rate. In future work we plan to investigate further the specific features learned by our model through feature activation analysis. We also plan to train and compare the performance of the model on other affective computing datasets for emotion. Finally, it would benefit the affective computing community greatly if we can develop optimal representations that can enable the aggregation of data using different devices and dataset to build a single model that contains all the richest features for HRV analysis.

References

- [1] G. Acampora, D.J. Cook, P. Rashidi, and A.V. Vasilakos. A survey on ambient intelligence in healthcare. *Proceedings of the IEEE*, 101(12):2470–2494, Dec 2013.
- [2] Gari D Clifford and Lionel Tarassenko. Quantifying errors in spectral estimates of hrv due to beat replacement and resampling. *IEEE transactions on biomedical engineering*, 52(4):630–638, 2005.
- [3] T Jr Deeming. Fourier analysis with unequally-spaced data. *Astrophysics and Space Science*, 36(1):137–158, 1975.
- [4] Juan Lorenzo Hagad, Ken-ichi Fukui, and Masayuki Numao. A hierarchical model of authentic work stress using physiological signals and stress coping profiles. In 人工知能学会全国大会論文集 2016 年度人工知能学会全国大会 (第 30 回) 論文集, pages 3E34–3E34. 一般社団法人人工知能学会, 2016.
- [5] H. Iso, C. Date, A. Yamamoto, H. Toyoshima, N. Tanabe, S. Kikuchi, T. Kondo, Y. Watanabe, Y. Wada, T. Ishibashi, H. Suzuki, A. Koizumi, Y. Inaba, A. Tamakoshi, and Y. Ohno. Perceived mental stress and mortality from cardiovascular disease among Japanese men and women: the Japan Collaborative Cohort Study for Evaluation of Cancer Risk Sponsored by Monbusho (JACC Study). *Circulation*, 106(10):1229–1236, Sep 2002.
- [6] Markad V Kamath, Dhanjoo N Ghista, Ernest L Fallen, David Fitchett, Donald Miller, and Robert McKelvie. Heart rate variability power spectrogram as a potential noninvasive signature of cardiac regulatory system response, mechanisms, and disorders. *Heart and vessels*, 3(1):33–41, 1987.
- [7] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry investigation*, 15(3):235, 2018.
- [8] E. Kowata, A. Hozawa, M. Kakizaki, Y. Tomata, M. Nagai, Y. Sugawara, S. Kuriyama, and I. Tsuji. [Perceived stress and cardiovascular disease mortality. The Ohsaki Cohort Study]. *Nihon Koshu Eisei Zasshi*, 59(2):82–91, Feb 2012.
- [9] M. Marcus, M.T. Yasamy, M. van Ommeren, and D. Chisholm. Depression: A global public health concern. Technical report, World Health Organization, 2012.
- [10] Paolo Melillo, Marcello Bracale, and Leandro Pecchia. Nonlinear heart rate variability features for real-life stress detection. case study: students under stress due to university examination. *BioMedical Engineering OnLine*, 10:96+, November 2011.
- [11] G.B. Moody. Spectral analysis of heart rate without resampling. In *Computers in Cardiology 1993, Proceedings.*, pages 715–718, Sep 1993.
- [12] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*, 2017.
- [13] Jeffrey D Scargle. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.
- [14] Julian F Thayer, Fredrik Åhs, Mats Fredrikson, John J Sollers III, and Tor D Wager. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, 36(2):747–756, 2012.
- [15] Panagiotis Tzirakis, George Trigeorgis, Mihalis A Nicolaou, Björn W Schuller, and Stefanos Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1301–1309, 2017.
- [16] Jacob T VanderPlas. Understanding the lomb–scargle periodogram. *The Astrophysical Journal Supplement Series*, 236(1):16, 2018.
- [17] Peter Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.
- [18] Xinglei Zhu, Gerald T Beauregard, and Lonce L Wyse. Real-time signal estimation from modified short-time fourier transform magnitude spectra. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5):1645–1653, 2007.