Machine Learning-Based Identification of Burnt-out Individuals using ECG data

Arya Sinha - Mohammed Kaif - Amolika Bansal
ML Project Proposal
INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY, DELHI
arya20498@iiitd.ac.in kaif20084@iiitd.ac.in amolika20424@iiitd.ac.in

Abstract

Burnout is a state of emotional, mental, and physical exhaustion caused by prolonged exposure to high levels of stress. The early identification of burnout is crucial in preventing long-term physical and mental health problems. A dataset of 202 individuals with three ECG recordings and Mini-Z 1.0 stress questionnaire responses was collected at the Department of Cardiology in G.B. Pant Hospital in Delhi. This study proposes the use of machine learning algorithms to classify individuals as either burnt-out or not based on their ECG data. The results of this study will help in the early identification and prevention of burnout-related health problems.

Keywords ECG, Mental State, Burn-out, Therapy, Classification, Supervised Learning

1 Introduction

Burnout is a widespread problem in today's fast-paced and demanding society, and it can have severe consequences on an individual's physical and mental health. It is often associated with various health problems, including heart diseases. The electrocardiogram (ECG) is a non-invasive diagnostic tool used to assess the heart's electrical activity and consists of 12 leads - six chest leads and six limb leads - recorded from electrodes on the body's surface. Anomalies in ECG samples have been shown to be predictive of both short- and long-term death. Therefore, early and accurate identification of cardiac ECG irregularities can help prevent burnout-related health problems. However, manual ECG interpretation is time-consuming and requires highly skilled professionals. Therefore, automated prediction of burnout using raw ECG data has been rapidly advancing over the past few years. In this study, we propose the use of machine learning algorithms to classify individuals as either burnt-out or not based on their ECG data. The goal is to develop a model that can identify burnout early so that the necessary interventions can be initiated to prevent long-term negative consequences. The study will utilize a dataset of 202 individuals with three ECG recordings and Mini-Z 1.0 stress questionnaire responses collected at the Department of Cardiology in G.B. Pant Hospital in Delhi. The results of this study could have significant implications for the early detection and prevention of burnout-related health problems.

1.1 Literature review

- The traditional processing pipeline for ECG signals has been subject to several improvements. One such advancement was proposed by Mashrur et al. (2019) [3], who utilized the continuous wavelet transform (CWT) to convert single-lead ECG data into RGB images. This was sent as an input to AlexNet for analysis. In a similar vein, Jun et al. (2018) [2] transformed ECG data into a 2D greyscale image and used this image as an input for classification with AlexNet and VGGNet.
- A research study created an automatic system for detecting mental stress using ECG signals and analyzed with ML classifiers, including decision tree, random forest, Naive Bayes. The most effective model had a `94.1%` accuracy rate[1].
- Another study by Zhang et al. (2014) proposed a classification method that utilized kernel SVM and Genetic Algorithm (GA). The approach had three primary modules: lead-fall detection, feature

extraction, and classification. The method achieved a true positive detection rate of 92% with 5.68% false positives and 94% classification accuracy using a training dataset with 1000 records [4].

- Zhou et al. (2021) conducted research in three phases. In the first phase, they extracted HRV features from ECG signals. In the second phase, they used Gaussian mixture model (GMM) based on HRV features to determine mental state. To reduce the stress classification coefficient (SCC), they applied a clustering approach to process HRV features. They used a Conventional Neural Network to extract features and combined manual and automatic features before passing them to SVM. The average recognition rate was around 95% [5].

1.2 Objective

- **Feature extraction:** Extracting relevant features from the preprocessed ECG data that can help differentiate between burnt-out and non-burnt-out individuals. Some possible features could be heart rate variability, QT interval, P wave duration, and T wave amplitude.
- **Building a classification model:** Developing a machine learning model that can classify individuals into burnt-out and non-burnt-out categories based on their extracted ECG features. This could involve experimenting with different classification algorithms, hyperparameters, and model architectures to achieve the best possible accuracy.
- **Model evaluation and validation:** Evaluating the performance of the developed model using suitable metrics such as accuracy, precision, recall, and F1-score, and validating it on an independent dataset to ensure its generalizability.

1.3 Scope

The identification of burnt-out individuals, using Machine Learning techniques on ECG data, has the potential to open up a new channel for the early detection and intervention of burnout, which is a serious problem in today's society. Machine learning algorithms can be trained to recognise patterns in ECG signals and detect changes related to burnout. ECG data can reveal information on the autonomic nervous system's activity, which is in charge of regulating the body's stress response. Machine learning algorithms can detect changes in heart rate variability, which is a sign of stress, by examining ECG data.
The application of machine learning-based identification of burnt-out individuals using ECG data has a broad scope. It can be employed in a variety of industries, including healthcare, education, and corporate settings. Technology in healthcare can be used to detect burnout in healthcare workers, lowering the risk of medical errors and increasing patient outcomes. Technology can be utilized in the education industry to identify burnt-out pupils and provide timely treatments to improve their mental and emotional wellbeing. In the business environment, technology can be utilized to recognise employee burnout, lowering absenteeism and increasing productivity.

1.4 Impact

The results of this study could have significant implications for the early detection and prevention of burnout-related health problems.
The diagnosis of burnt-out individuals using machine learning techniques on ECG data has the potential to change the way we address burnout in our society. This device can enable early identification and intervention for people who are at risk of burnout, enhancing their mental and emotional well-being and averting severe burnout. Detecting and managing burnout early on can result in increased productivity and lower healthcare expenses in the long run. This technology can improve people's quality of life by preventing health problems and boosting their general well-being.
In conclusion, this can have a tremendous positive impact on our society, increasing individuals' mental and emotional health and enhancing productivity and overall well-being.

2 Materials and Methods

2.1 Dataset

The experiment was conducted at the Department of Cardiology at G.B. Pant Hospital in Delhi and collected ECG data. Raw ECG data was saved with 12 leads, 500 Hz, and 10 seconds. Two hundred two-people were selected to participate in the study. Participants visited the study center three times (visit 1, visit 2, visit 3) when their ECG was recorded. All participants completed a study proforma that included socio-demographic information, clinical details, and the Mini-Z 1.0 stress questionnaire during each visit. On the Mini-Z questionnaire, the participant's response to question 3 indicates if the person is feeling burnout or not.

2.2 Methodology

- **Preprocessing:** Preprocess the dataset to prepare it for analysis. This may involve cleaning the data, removing outliers, and normalizing the ECG readings.
- **Feature selection/extraction:** Select relevant features from the dataset. This could include ECG's lead I, II, V1, V2, V3, V4, V5, and V6. Following this, HRV features can be extracted.
- **Model training:** Train a machine learning classification model on the dataset using supervised learning algorithms, logistic regression, KNN, ANN to classify the mental state i.e. burnout/stressed or not stressed of an individual.
- **Model evaluation:** Once the model is trained, evaluate its performance by testing it on new data that the model hasn't seen before.

2.3 Evaluation Metric

We will be computing the **Confusion Matrix** - a table that shows the number of true positives, true negatives, false positives, and false negatives.

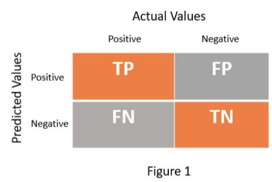

Figure 1: Confusion Matrix

From the confusion matrix - Accuracy, Sensitivity and Specificity is evaluated using the following equations:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TP + FP}$$

F1 score: the harmonic mean of precision and recall. It provides a balanced measure of both precision and recall.

ROC AUC: the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate. It measures the ability of the model to distinguish between positive and negative samples.

2.4 Novelty

Our aim is to classify burnt-out individuals based on ECG data, which has not been extensively explored in previous research. While several studies have investigated the relationship between stress and cardiovascular disease, few have used ECG data to identify burnout specifically. Also, the dataset that we will be working on is a real world dataset and thus no previous work has been done using this. The data contains two groups, group 1 is performing Satyam meditation while group 2 is performing control mediation. We can inspect the differences in stress reduction caused by the two techniques.

3 Plan and Tentative Timeline

1. `[1 week]` Data pre-processing
2. `[2 weeks]` Feature Extraction and EDA
3. `[3 weeks]` Apply different methods & models to classify - Logistic regression, K-NN, ANN, etc.
4. `[2 weeks]` Report and presentation

4 Distribution of work

- **Arya Sinha** - Data pre-processing, Logistic regression, ANN, CNN, SVM
- **Mohammed Kaif** - Data pre-processing, KNN, SVM, Naive Bayes, Presentation
- **Amolika Bansal** - EDA, Logistic regression, ANN, Report Formation

References

[1] Md Belal Bin Heyat, Faijan Akhtar, Syed Jafar Abbas, Mohammed Al-Sarem, Abdulrahman Alqarafi, Antony Stalin, Rashid Abbasi, Abdullah Y. Muaad, Dakun Lai, and Kaishun Wu. Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal. *Biosensors*, 12:427, 06 2022.

[2] Tae Joon Jun, Hoang Minh Nguyen, Daeyoun Kang, Dohyeun Kim, Daeyoung Kim, and Young-Hak Kim. Ecg arrhythmia classification using a 2-d convolutional neural network. *arXiv:1804.06812 [cs]*, 04 2018.

[3] Fazla Rabbi Mashrur, Amit Dutta Roy, and Dabasish Kumar Saha. Automatic identification of arrhythmia from ecg using alexnet convolutional neural network. *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, 12 2019.

[4] Ya-tao Zhang, Cheng-yu Liu, Shou-shui Wei, Chang-zhi Wei, and Fei-fei Liu. Ecg quality assessment based on a kernel support vector machine and genetic algorithm with a feature matrix. *Journal of Zhejiang University SCIENCE C*, 15:564–573, 07 2014.

[5] Ruishi Zhou, Chenshuo Wang, Pengfei Zhang, Xianxiang Chen, Lidong Du, Peng Wang, Zhan Zhao, Mingyan Du, and Zhen Fang. Ecg-based biometric under different psychological stress states. *Computer Methods and Programs in Biomedicine*, 202:106005, 04 2021.