

Design and Implementation of a Single-Lead Chest Strap ECG Recorder for Stress Classification using Lightweight Machine Learning Methods

BEng Final Year Project Report

William Powell

Integrated Mechanical and Electrical Engineering

Student No: 199126158

Supervisor: Dingguo Zhang

Academic Year: 2022/2023

Word Count: 12260

Referencing Type: IEEE

May 2023

Abstract - In this report, an affordable, chest strap Electrocardiogram (ECG) device that detects stress using lightweight Machine Learning (ML) methods in a real-life environment, was designed, built and evaluated. To the best of our knowledge, no previous studies have demonstrated the feasibility of stress detection using an ECG device of this kind that does not rely on wet and/or disposable electrodes. Firstly, a comprehensive review of stress biomarkers and their associated features were conducted, and demonstrated ECG to be a versatile candidate for determining the level of sympathetic activation. Heart Rate Variability (HRV), waveform intervals, and ECG-Derived Respiration (EDR) features in the time, frequency and nonlinear domains were selected for feature extraction. A small number of the most effective and lightweight supervised ML models were chosen from other literature findings. A demonstrative study was conducted, utilizing data collected from three male participants aged 21-22 years who wore the device while studying. The study showed that the device could capture ECG with minimal noise and artifacts despite the lack of a Right Leg Drive (RLD) electrode. The location of the electrodes showed to be an important issue that must be resolved to allow for effective detection of the ECG's Q wave, which is required for numerous features involving ventricular repolarization. Random Forest demonstrated to be the most effective model for both binary (low and high) and three-level (low, medium and high) stress level classification offering a good compromise between resource savings and accuracy. The One-dimensional (1D) Convolutional Neural Network (CNN) showed promise as its lightweight structure allowed for comparable runtime to traditional models and benefits from automatic feature extraction and ultra-short window size (3s). However, insufficient training data was used to confirm its effectiveness in the context of this study. Window size showed to severely affect the accuracy in all models with a drastic decrease from 30 to 20s, whilst sampling rate showed only a minor reduction in performance even down to 100Hz. The study shows that the designed chest strap ECG device could be an effective tool for detecting stress in the context of daily use and has the potential to outperform other devices, such as PPG wrist wearables, in terms of accuracy and reliability.

Acknowledgements

I would like to express my sincere gratitude to Dr. Dingguo Zhang for his support and guidance throughout this project. I would also like to thank Dr. Frank Ihmig for generously providing me with the subjective ratings data from his paper and for his enthusiasm in my work. Additionally, I am grateful to Dr. Konstantinos Tzevelekakis for his support in expanding on his implementation and architecture of the 1D-CNN as presented in his paper. Their contributions have been invaluable to the completion of this project.

Contents

1	Introduction	1
1.1	Background	1
1.2	Motivation, Rationale and Objectives	1
1.3	Biomarkers and Biosignals of Stress	2
1.4	ECG Signatures of Stress	3
1.5	Supervised Machine Learning Techniques for Stress Classification	5
2	Device Design	9
2.1	Specifications	9
2.2	System Overview	10
2.3	Hardware	10
2.3.1	Signal Conditioner	10
2.3.2	Analog-to-Digital Converter	11
2.3.3	Electrodes	12
2.3.4	Other Components	12
2.3.5	PCB Design	13
2.4	Firmware	13
3	Experimental Method and Device Results	15
3.1	Protocol and Data Acquisition	15
3.2	Methodology Overview	16
3.3	Signal Analysis and Preprocessing	17
3.3.1	Signal Quality Analysis	17
3.3.2	Filtering	18
3.3.3	Data Preprocessing	19
3.4	Feature Extraction	20
4	Classifier Implementation and Results	25
4.1	Naive Bayes	25
4.2	Support Vector Machine	26
4.3	Random Forest	28
4.4	1D Convolutional Neural Network	29
4.5	Model Comparison	31
4.6	Feature Selection	32
4.7	Optimisation	35
5	Discussion	37
5.1	Device Performance	37
5.1.1	Signal Quality	37
5.1.2	Signal Reliability	37
5.1.3	Feature Extraction	37
5.2	Classification Performance	38
5.2.1	Model Comparison	38
5.2.2	Feature Set Comparison	38
6	Limitations and Future Work	39
7	Project Review and Conclusions	40
A	Appendix	49

List of Abbreviations

1D	One-dimensional
ADC	Analogue to Digital Converter
ANN	Artificial Neural Network
CMRR	Common-Mode Rejection Ratio
CNN	Convolutional Neural Network
DTs	Decision Trees
ECG	Electrocardiogram
EDA	Electrodermal Activity
EDR	ECG-Derived Respiration
EEG	Electroencephalogram
GVS	Galvanic Vestibular Stimulation
HR	Heart Rate
HRV	Heart Rate Variability
IMU	Inertial Measurement Unit
KDE	Kernel Density Estimate
ML	Machine Learning
MLP	Multilayer Perceptron Neural Network
OSR	Oversampling Ratio
PCB	Printed Circuit Board
PNS	Parasympathetic Nervous System
PPG	Photoplethysmography
PSD	Power Spectrum Density
RCT	Randomised Controlled Trial
RESP	Respiration Signal
RLD	Right Leg Drive
RSA	Respiratory Sinus Arrhythmia
SMOTE	Synthetic Minority Oversampling Technique
SNS	Sympathetic Nervous System
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TEB	Thoracic Electrical Bioimpedance
UDP	User Datagram Protocol
VRV	Ventricular Repolarization Variability

1 Introduction

1.1 Background

ECG has demonstrated to be a versatile measurement to diagnose heart conditions such as arrhythmia, heart damage and the most common cause of death - cardiovascular disease [1, 2, 3]. In addition to detecting abnormalities, ECG can evaluate more subtle variances such as athlete performance or stress levels [4, 5]. Excessive stress accumulation can cause burn-out [6, 7] that currently affects one in five of the UK workforce [8]. Long-term effects include suppression in the immune system, and increase the likelihood of chronic diseases such as high blood pressure, heart disease, and cancer [9]. According to the World Health Organization, there has been a 25% surge in anxiety and depression cases globally due to the COVID-19 outbreak [10], exposing the current lack of tools to diagnose and treat mental health issues. Automated stress detection has shown to alleviate symptoms of mental health disorders [11], and synthesising this with stress mitigation techniques such as Galvanic Vestibular Stimulation (GVS) [12], shows great potential for diminishing this global problem.

1.2 Motivation, Rationale and Objectives

This project has been motivated by BrainPatch, a Neurotechnology Company, who aim to combat stress through GVS. This has been exemplified by a number of studies but particularly Pasquier et al. 2019 [12], which demonstrates that GVS is able to modulate anxiety and corroborates the involvement of the vestibular system in the emotional process. The use of a wearable device, capable of recording a user's ECG, was proposed as a suitable method to detect the effectiveness of this stress relief.

The aim of the project is to demonstrate that a low-cost, chest strap ECG system can be used to detect stress in everyday life. To the best of our knowledge, no previous studies have demonstrated the feasibility of stress detection using an ECG device of this kind that does not rely on wet and/or disposable electrodes. Commercial chest strap heart rate monitors have gained traction in the fitness market, indicating that this form factor is comfortable and practical for long-term use and eliminates the need for wet/disposable electrodes. As explained in Section 2.1, there are currently few commercial chest strap ECG devices that offer sufficient specification in terms of recording duration, signal quality, motion detection capability, and affordability. Hence, this report will detail the design of a non-clinical wearable ECG device that offers the specifications mentioned. The device's efficacy and suitability for stress classification will then be evaluated through the use of several machine learning techniques for binary and three-level classification.

Objectives:

1. **Research Aims:** Research and implement ML models that detect key stress indicators.
 - (a) Evaluate and determine most suitable biomarkers and thus features for stress classification.
 - (b) Select suitable ML models from the literature that are lightweight and demonstrate high accuracy for stress classification.
2. **Design Aims:** Design and implement an affordable ECG recorder that is suitable for stress detection.
 - (a) Minimise cost - the device must be comparable to that of a smartwatch to enable access of stress detection to a larger population.
 - (b) Achieve a high fidelity signal for identification of key components of the ECG, see Figure 2. Crucially, the peaks of the ECG must be accurately and reliably identified by automated peak detection software.
 - (c) Mitigate noise and minimising common ECG artifacts, such as movement, muscular and baseline drift, using analogue and digital signal processing.
 - (d) Integrate an Inertial Measurement Unit (IMU) to the recorder to measure movement, which can be accounted for during classification.

- (e) Achieve a high bit-rate wireless system capable of streaming high fidelity ECG recording (16 bits at 1kHz) in real time.
3. **Experimental Aims:** Evaluate effectiveness of decoding stress levels using the model(s) developed.
- (a) Obtain data using the device in a demonstrative study and evaluate against pre-existing stress labeled ECG database.
 - (b) Optimise accuracy, computational efficiency, and memory of models through hyperparameter tuning, number of features and window length.
 - (c) Determine the minimum sampling rate of the device before classification accuracy is significantly impaired.
 - (d) Determine effectiveness and limitations of the models chosen such as the feasibility of on-chip, low computational classification.

1.3 Biomarkers and Biosignals of Stress

As described in Rochette et al. [13], stress was first defined by Hans Selye as “the nonspecific response of the body to any demand made upon it” [14] and has since had varying definitions. Selye’s clear emphasis on the biological aspect, neglects the psychological side involved. Therefore for the context of this report, unless otherwise stated, stress will be defined more broadly as “a state of mental strain caused by external or internal factors which result in an undesirable physiological or psychological response”. This definition does not refer to chronic or prolonged stress, neither does it necessarily require a clear stressor like with active mental stress.

The biological response to stress is controlled by the hypothalamic-pituitary-adrenal and sympathetic-adrenal-medullary axes of the Sympathetic Nervous System (SNS), and are responsible for regulating the release of hormones such as cortisol, epinephrine, norepinephrine, and alpha-amylase [15], see Figure 1. These hormones increase heart and breathing rates, suppress the immune system, and prioritize energy towards vital physiological functions, preparing the body to survive a perceived threat [16, 17]. Conversely, the activation of the Parasympathetic Nervous System (PNS) pathway through stimulation of the vagus nerve indicates a mental relaxed state [18].

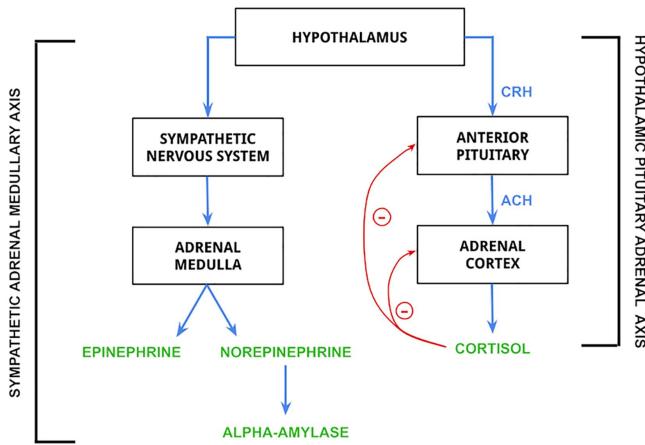


Figure 1: Mechanism of release for epinephrine, norepinephrine, alpha-amylase, and cortisol from the hypothalamus [19].

It has been contended that identifying biomarkers like cortisol secretion would provide more accurate indicators of stress compared to biosignals, since biosignals solely reflect the physiological outcomes stemming from the release of stress biomarkers rather than the underlying cause [19]. Clinical devices have been able to detect cortisol through electrochemical sensors, however, there has been little

commercial traction for stress biomarker detection. Instead, non-invasive biosignals are more commonly used to detect these changes to the nervous system and include Electrodermal Activity (EDA), Thoracic Electrical Bioimpedance (TEB), Photoplethysmography (PPG), ECG, and Electroencephalogram (EEG) [5, 20].

To detect stress using EEG, electrodes are typically placed on the prefrontal cortex and thus requires an EEG cap which is not suitable for daily use [21]. TEB requires numerous electrodes around the chest and is therefore also not suitable. EDA is a technique which measures the electrical conductance of the skin, which is affected by the activity of sweat glands [22]. This method is a popular technique since the sensor can be mounted on a wristwatch [23]. From a review of the literature, previous studies demonstrate that EDA and ECG signals are two of the most common signals for detecting stress that are commercially viable. BrainPatch have chosen ECG, with the aim of only using two electrodes composed as a chest strap, primarily since it offers the least visible and most comfortable solution.

1.4 ECG Signatures of Stress

ECG is a method that monitors the electrical signal generated by the heart during a cardiac cycle against time. During depolarization and re-polarisation of the myocardial fibers of the heart, a potential difference is created that can be recorded by electrodes located near the heart [24]. A typical ECG trace consists of a P wave, QRS complex, T wave, and U wave.

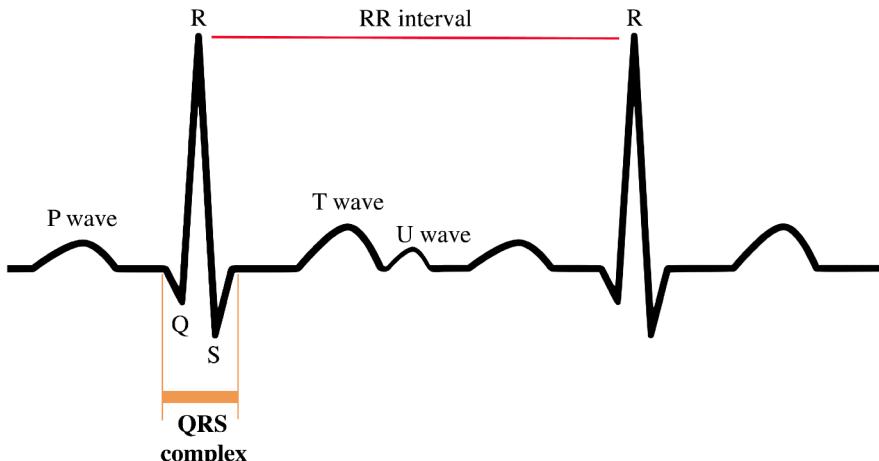


Figure 2: Ideal ECG trace of a normal sinus rhythm for a human heart in Lead II configuration [25].

HRV: The QRS complex, and in particular the R peaks, are the most common markers for identifying stress. The R-R interval is the time difference between two consecutive R peaks, usually measured in milliseconds. Heart Rate (HR) is the average R-R interval per minute, and is thus offers a lower resolution indicator to changes to the SNS/PNS. An excellent indicator of stress is HRV - the variation of R-R intervals [24]. A high HRV indicates a normal balance between SNS and PNS activation, while a low HRV indicates high mental stress due to the activation of the SNS and inhibition of the PNS. Through a meta-analysis of 12 papers, Castaldo et al. [26] demonstrated that HRV was significantly altered in the time, frequency and nonlinear domains. These significant features are illustrated in Figure 6.

Parak et al. has shown that HRV measurements recorded through a wrist-worn PPG device, an optical heart monitor, yields comparable accuracy and reliability to HRV obtained through chest strap ECG [27]. It was also found that participants prefer wrist and arm-worn devices in terms of aesthetics, wearability, and comfort over chest-worn devices since it does not require electrodes to be placed on the body. However, as discussed in this study, ECG provides a richer signal to PPG since it can measure additional stress-related features, thereby improving its classification accuracy over HRV measurement alone.

VRV: In addition to HRV analysis, ECG provides other stress indicators based on detecting changes to SNS/PNS activation. Ventricular Repolarization Variability (VRV), which can be obtained from QT wave interval has a direct relation between HRV and Respiratory Sinus Arrhythmia (RSA), and is hence another indicator of SNS activation [28, 29]. VRV can be measured by the variability in the duration of the QT interval, which is the time between the start of the QRS complex and the end of the T wave [30]. Andrassy et al. [31] determined that sudden mental stress increases QT prolongation, however ascertained the important distinction that the QT interval must be corrected for R-R intervals, due to conflicting studies that did not take this into account [32, 33]. Nonetheless, QT intervals provide a valuable indication of stress in healthy individuals and this metric is not available through optical heart rate sensors like PPG.

EDR: In addition, from the use of R wave amplitudes, a respiration signal can be derived which is similarly affected by stress. RSA is the cyclic variations in heart rate due to respiration [34] and is commonly used to estimate autonomic cardiovascular control [35]. EDR has shown to be an excellent correspondence to respiration signals generated by a respiration belt in both the time and frequency domain [36, 37]. Furthermore, Imam et al. [29] evaluated the respiration signals recorded with an elastic Hall Effect sensor that measured chest cavity expansion of a publicly available stress labeled dataset, DriveDB [38] and compared it to that of EDR.

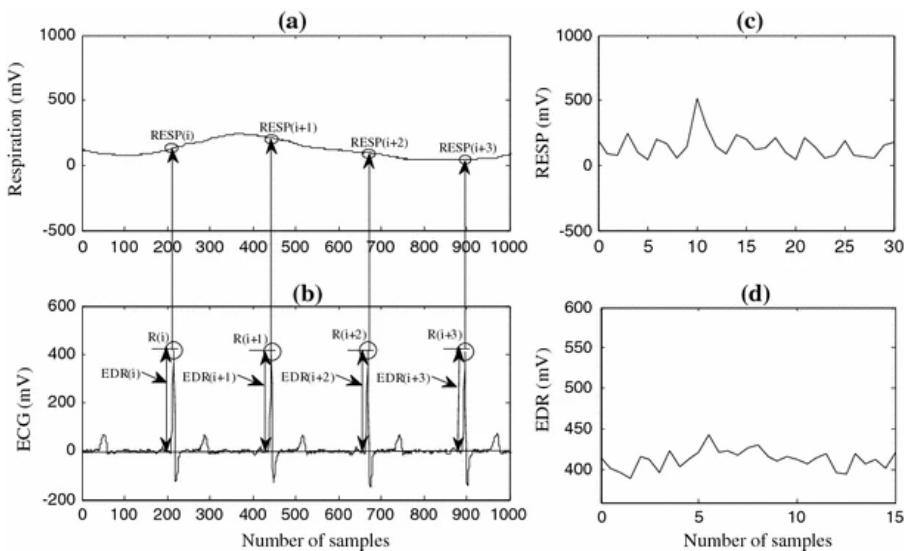


Figure 3: Derivation of Respiration Signal (RESP) and EDR time series from the thoracic belt respiration and baseline-corrected ECG signal [29].

After baseline wandering was corrected by a median filter, the QRS wave was detected from the ECG (b) using the Pan Tompkins algorithm and the R wave detected from the local maximum amplitude. From this, the R peak was used to obtain a maximum value of the respiration signal (a) to generate the RESP signal (c). The amplitudes of the R peaks were used to generate the EDR signal (d). Using recordings from 17 subjects, it was shown that QT–RR–EDR models from short-term ECG resulted in significantly better fit for binary stress classification than that found from the QT–RR model and showed almost the same level of predictability as that of the models derived using respiration (QT–RR–RESP). It was concluded that EDR can be used as a surrogate of respiration for modeling ventricular repolarization dynamics in stressed situations, thus further demonstrating the advantages of ECG devices over typical wrist wearables that only measure R-R intervals.

1.5 Supervised Machine Learning Techniques for Stress Classification

From the Advanced Signal Processing in Wearable Sensors for Health Monitoring Journal [39], there are a diverse range of ML algorithms used with ECG signal processing which vary in complexity, intelligence, and generalizability. Stated in Krittawong et al. [40], the most common ML algorithms used for ECG analysis include: Naive Bayes [41], Support Vector Machine (SVM) [42] [43], Random Forest [44], and KNN [45] in addition to Artificial Neural Network (ANN)s, such as MLP [46] and CNN [47]. The models that have demonstrated both lightweight architecture and high accuracy for stress classification are detailed below.

Naive Bayes

Keshan et al. [48] evaluated different ML algorithms using DriveDB [38], a well-studied stress labelled ECG database. The features selected involved average duration and intervals between Q, R and S waves in the time domain only. Through the algorithms tested, Naive Bayes showed to be most accurate for binary classification and achieved 100% accuracy when trained on restful ECG for an individual participant by using average heart-rate or average RR, QQ or SS durations. Naive Bayes is arguably the most simplistic algorithm for ECG classification, since it favours few features, small datasets and is best suited for low-computational, binary classification. It also does not suffer when the dataset contains class imbalance since this can be factored into the posterior probabilities. Hence, Naive Bayes will be used to evaluate the ECG recorder in most simplistic form.

$$p(C|A) = \frac{p(A|C) \cdot p(C)}{p(A)} \quad (1)$$

Bayes' theorem gives the relationship between the probabilities of C and A, $P(C)$ and $P(A)$, and the conditional probabilities of C given A and A given C, namely $P(C|A)$ and $P(A|C)$. Bayes' theorem is efficient since $P(C|A)$ is often much more computationally expensive than $P(A|C)$ and $P(C)$ [49]. Naive Bayes uses Bayes' theorem with the "naive" assumption of independence between every pair of features. Despite this oversimplification, Naive Bayes is often used due to its simplicity and efficiency. To improve the accuracy of the Naive Bayes classifier, careful selection of the input data distribution is required, which may include Gaussian, Multinomial, and Bernoulli distributions.

Support Vector Machine

To achieve generalizability between users, in addition to variations in real-life environments that affect ECG, more sophisticated algorithms and features should be encouraged to differentiate the signal's characteristics. Krittawong et al. stated that ANNs and SVMs are the two most popular classifiers for cardiovascular ECG classification due to their ability to classify nonlinear relations and compatibility with larger datasets. Given that features often do not align well with a linear decision boundary, it is advisable to employ nonlinear classifiers or feature crosses, where nonlinear feature datasets can be transformed into linear models by multiplying two features together [50]. SVM employs nonlinear decision boundaries and is often less computational intensive to neural networks, thus could offer a good compromise between performance and complexity [51].

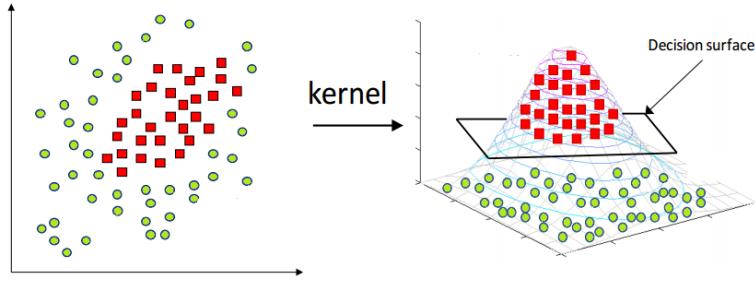


Figure 4: Kernel data transformation (Radial Basis Function) used to classify nonlinear features by converting to a higher dimensional space [52].

The main objective of SVM is to find a hyperplane (decision boundary) that can maximize the margin between classes, leading to the best possible separation of the input data. When the classes are not linearly separable, kernel functions can be used to map the observations into a higher-dimensional space, where the hypersurface (decision surface) can linearly separate the classes [51].

Random Forests

Ihmig et al. [53], demonstrated through a Randomised Controlled Trial (RCT) of 56 arachnophobic individuals, that 3 levels of stress could be classified with up to 74.4% accuracy using ECG, EDA and RESP biosignals with only a 10 second window length and six features used. 73.5% was achieved using only three features (two for ECG and one for EDA), however it is unclear to what extent the EDA feature improved the accuracy. From the evaluation of several traditional ML classifiers, the Random Forest algorithm showed to have the best performance. Due to the relatively low computational requirements of Random Forest, in addition to few features and a short window length, real-time on-chip detection may be possible using this algorithm, hence it shall be investigated.

Random forest is a type of ensemble classifier that involve the creation of multiple Decision Trees (DTs). DTs comprise of nodes and branches where each node represents a test on a particular feature, and each branch represent the outcome of that feature. The decision of multiple branches by summation yields the classification [40]. DTs suffer from high variance - whereby the splitting of the training and testing data will produce very different results. To combat this, Bagged Trees (Bootstrap Aggregated Trees) a type of ensemble classifier, use several DTs and conduct bootstrap aggregation, where repeated sampling with replacement and aggregation of the results is conducted to reduce variance in the model. Random forest, an extension to Bagged Trees, introduce randomness, by only using a subset of features randomly selected and the best split feature from the subset is used to split each node, thereby further distancing from the variant nature of DTs and reduce over-fitting.

Convolutional Neural Networks

Unlike the traditional ML algorithms mentioned above, deep learning neural networks are capable of performing automatic feature extraction without human intervention [54]. This can be useful since extracting features such as HRV statistical parameters often require large libraries and computation is not usually efficient. Keshan et al. identified that Multilayer Perceptron Neural Network (MLP) neural networks resulted in the best classification for 3-level classification, however this required a large window size of 5 minutes, which would thus require large memory and computational efficiency, which is not suited for on-chip classification. However, many studies have shown that 1D CNNs can achieve accurate results in various ECG classification tasks, even when using a small window [47, 55, 56, 57, 58, 59]. For example, Tzevelekakis et al. [47], demonstrated how deep CNNs could be used to determine stress levels using ultra-short (3s), raw ECG Signals, which demonstrate to change in morphology during stress [60]. Using 1D CNNs, an accuracy of 83.55% and 98.77% was achieved for two and three-level stress classification respectively. Two pre-recorded publicly available datasets were used - DriveDB and Arachnophobia [53] which contained annotated ECG signals for 3-level stress

classification. This study demonstrates that through short windowed samples, lightweight ANNs could be used for accurate classification which have a low memory footprint and could be utilised for on-chip detection. Due to the close alignment of this study, 1D CNN will be investigated.

ANNs consist of layers containing neurons similar to the biological brain. Each neuron has a set of inputs and a specific weight that are used to computes a function. The neuron's weight is updated during training and takes a linear combination of weighted input and usually applies a nonlinear activation function, most commonly sigmoid, tanh or relu, on the aggregated sum [51].

CNNs are a type of ANN that uses convolutional layers containing filters or kernels. 1D CNNs are usually used for time-series data such as raw ECG, where the kernel is a one-dimensional sequence [61]. The convolutional operation is denoted as:

$$y_{mn} = f\left(\sum_{J=0}^{J-1} \sum_{I=0}^{I-1} x_{m+i,n+j} \omega_{ij} + b\right) \quad (2)$$

In this context, x denotes the matrix that undergoes the convolution operation, and y is the resulting output. I and J refer to the dimensions of the convolution kernel, while ω represents the weights, b denotes a bias, and f represents the activation function.

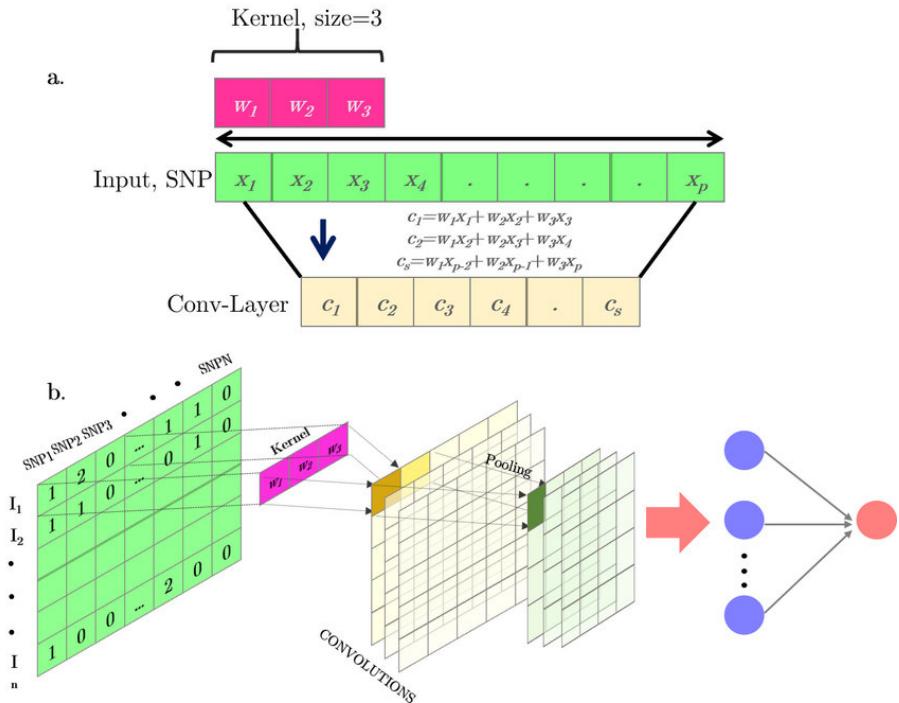


Figure 5: (a) Simple scheme of a 1D convolutional operation. (b) Full representation of a 1D convolutional neural network for an input matrix. Where yellow represents the outputs of the convolutional layer, green represents the pooling layers, and red is the output generated by a standard MLP [62].

The filter shifts across a sliding window of the input sequence, or feature vector, with a set stride, applying convolution to each input which results in an output sequence, or feature map. Pooling layers are used to downsample the output of the convolutional layers by reducing the spatial dimension of the feature maps from the maximum or average value of each subsequence. Controlling the size of the feature map is important and can be calculated using the formula below.

$$o = \left[\frac{i + 2p - k}{s} \right] + 1 \quad (3)$$

Where o is the output size of the feature map, i is the input size of the feature vector, p is the padding size, k is the kernel size and s is the stride length [63].

The size of the pooling window and the stride are hyperparameters that can be tuned to control the reduction in the temporal dimension [63]. A larger pooling window and stride result in more aggressive reduction, which can be useful in situations where the input sequence is very long and the model needs to capture high-level features, however due to complex and subtle features of raw ECG, this may result in loss of information. Padding can also be added to the feature vector to maintain or lengthen the feature map, thereby preventing loss of information from the edges of the signal.

2 Device Design

2.1 Specifications

As shown in Objective 2, the main design goal for this ECG device is to improve comfort without compromising signal quality. Thus, the device must have the following specifications:

1. **Comfort** - The device must be comfortable to wear all day, critically this means that only two dry electrodes should be used, see Section 2.3.3.
2. **Sampling Frequency and Resolution** - For accurate identification of the QRS complex using the most common algorithms, it has been shown that a minimum sampling frequency of 120Hz and bit-depth of 12-bits is sufficient [64]. Lowering the bit-depth more than this yields a low signal-to-quantization-noise ratio and would produce errors in beat detection algorithms. For frequency-domain HRV features, a sampling rate as low as 125Hz achieved almost identical classification to the original 2000Hz signal [65], utilising a 16-bit Analogue to Digital Converter (ADC). However, this conflicts with another studies stating that 250Hz is required for frequency domain [66] and for accurate readings of rapid ECG components such as R-R intervals [67]. Both studies agreed that there are negligible differences between 250Hz and 500Hz. Due to the benefits of implementing a high Oversampling Ratio (OSR), see Section 2.3.2, a high sampling ADC can be used in this investigation and can be later reduced to optimise classification speed and accuracy, especially for on-chip detection.
3. **Wireless Performance** - Clinical ECG devices are usually wired with digital isolation to improve signal quality and prevent wireless communication dropouts. Of course, during everyday use this is not desirable and occasional dropouts are tolerable. A wireless device allows for the raw signal to be transmitted, analysed, and if required, classified in real time. If the study demonstrates that on-chip classification is possible, wireless communication will only be required for alerting the user of the result and thus does not require high bandwidth.
4. **Price** - As demonstrated in Section 1.1, due to the prevalence of global stress and anxiety, the device must be low-cost to enable access to a larger population. Hence component price is an important consideration in the design.

The table below illustrates the few commercial ECG devices currently on the market that offer single-lead, wireless ECG.

Device	Electrode Type	IMU	Internal Storage	Wireless Protocol	Streaming	Recording Duration	Sampling Rate	Resolution	Price
Polar H10	Chest Strap	Yes	400h (RR only)	Bluetooth	RR and ECG	2h	130Hz	8 bit	£76
AliveCor KardiaMobile	Handheld	No	5m	Bluetooth	ECG	5m	300Hz	16 bit	£70
Wellue 24hr ECG Recorder	Adhesive / Strap	No	24h	Bluetooth	ECG	24h	Unknown	Unknown	£240
Seers mobiCARE-MC100	Adhesive	Yes	9h	Bluetooth	ECG	9	256 Hz	Unknown	N/A

Table 2: Specifications of popular, commercial, wireless, single-lead ECG devices.

KardiaMobile has shown to be effective for atrial fibrillation identification [68], however is not suitable for everyday use due to its small recording window. No research papers were visible for the Wellue device, perhaps due to the expense of the device. Due to the lack of IMU, stress classification would be unreliable. The Polar H10, has shown to perform well in numerous studies investigating HRV and HR recording during exercise and real-life environments [69, 70, 71, 72]. Only one study was found that utilised the device's ECG signal, but demonstrated effective identification of abnormalities such as atrial fibrillation [73]. This study did however require constant Bluetooth connectivity for ECG recording due to the device's small internal memory and therefore not suitable for everyday use.

2.2 System Overview

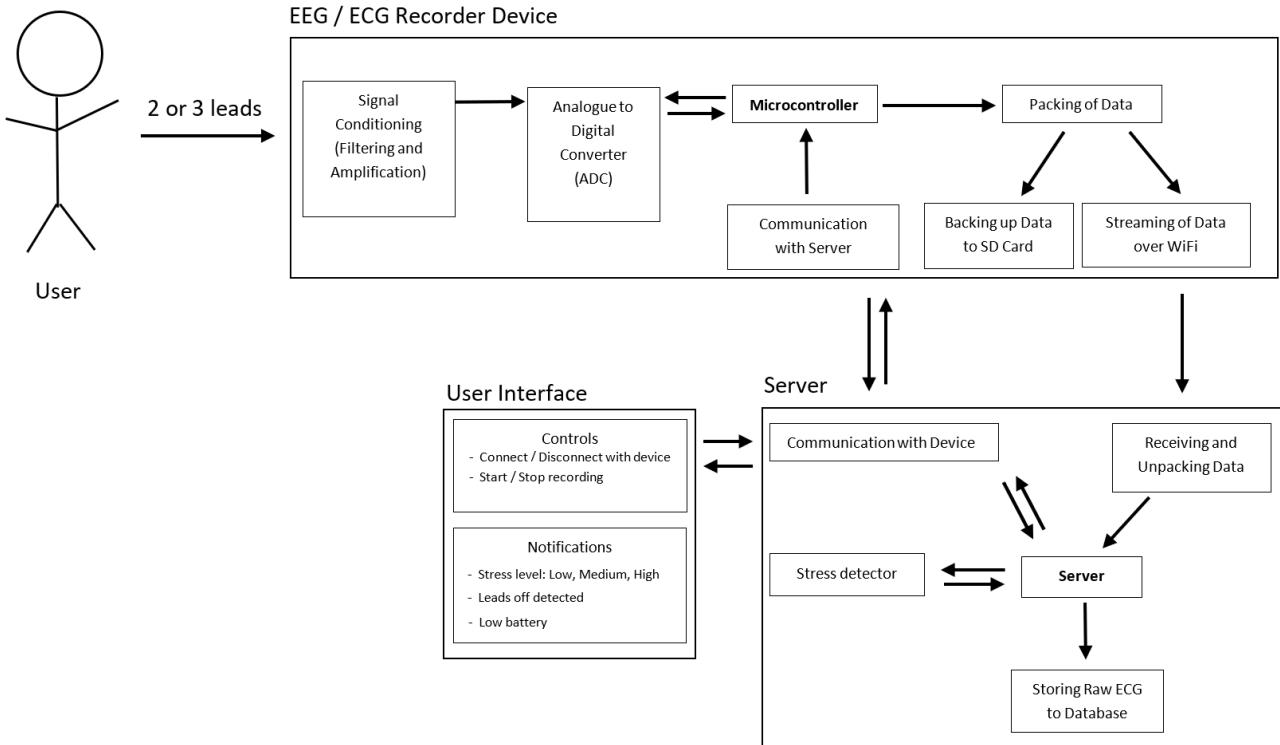


Figure 6: High-level system overview illustrating how a biosignal could be recorded, streamed and classified in real time.

The biosignal will first pass through the analogue front-end which filters and amplifies the signal through the signal conditioning unit from micro-volts to volts, preparing the signal for analogue to digital conversion. Once the signal is digitised, the microcontroller then packages the data and simultaneously stores to the SD card and streams over WiFi in real-time. The server then receives this data, where digital signal processing and stress classification can be conducted. To limit the scope of this report, the software for the server and user interface will not be discussed in this paper.

2.3 Hardware

2.3.1 Signal Conditioner

The signal conditioning device must be capable of filtering and amplification of extremely low amplitude biopotential signals, that are extremely susceptible to noise. To ensure this, the integrated circuit must have the following specifications:

- **Amplification** - The ECG signal generated has a high source impedance and their voltages range from 1-5mV [74]. If the input impedance of the signal conditioner's instrumentation amplifier is not sufficiently high, the signal source will be loaded, which will result in signal attenuation, distortion, and loss of signal quality. This source impedance is further magnified by the use of dry electrodes, chosen in Section 2.1, which are between 100-1000 times higher impedance than wet electrodes [75]. The bias current that flows through the input terminals of the amplifier must be kept low. When the bias current is high, it can create voltage offsets, which will create an artifact on the ECG signal [75]. It is important to mitigate clipping, where the signal surpasses the rail of the operational amplifier's (op-amp) output. This is typically achieved through a programmable gain, in addition to a rail-to-rail op-amp. which allows the output voltage to swing as close as possible to the power supply rails, while maintaining a linear response [76]. There must be sufficiently high headroom to prevent clipping when noise is induced in the system. This must be considered when setting the op-amp gain.

- **Filtering** - Typical ECG signals contain both high-pass and low-pass filters. The high-pass filter attenuates low-frequency components such as motion artifact, respiratory variation, and baseline wander, whilst the low-pass filter attenuates high-frequency muscle artifact and external interference [77]. Baseline wander arises from breathing, electrically charged electrodes, such as static, or subject movement which can alter the shape of ECG and in particular the ST segment [78]. This artifact may be pivotal in stress detection since the ST segment has shown to be a significant feature impacting the accuracy stress classification as shown by Goel et al. [79].
- **Noise Rejection** - ECG is vulnerable to common-mode interference from the 50/60-Hz mains. One technique to attenuate this noise is by utilising an additional 3rd electrode placed in a location away from the biopotential signal [77]. This electrode is then driven by an out-of-phase common-mode signal, called the RLD, thereby cancelling out the common-mode signal [80]. As discussed in Section 2.1, the aim is to not utilise this electrode to improve user comfort for long duration recording. Hence, the signal processing circuitry must have sufficiently high Common-Mode Rejection Ratio (CMRR) to mitigate the impact of noise. Mettingvanrijn et al. states that a biosignal amplifier should have a minimum CMRR of 80dB at 50Hz [81]. Power Supply Rejection is also important to eliminate noise from the power supply, as well as ensuring the noise at the output of the voltage regulator is low.

The signal conditioner chosen that meets the requirements above was the AD8232 - a fully integrated single-lead ECG front end, see datasheet. It features a sufficiently high common-mode rejection ratio of 80dB at 60Hz, rail-to-rail output, and programmable gain.

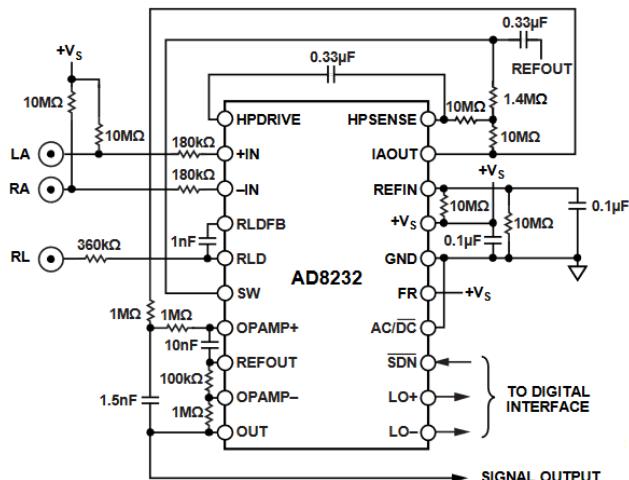


Figure 7: Signal Conditioning Circuitry that minimises signal distortion through the use of a RLD, 0.5 Hz two-pole high-pass and two-pole, 40 Hz, low-pass filter (obtained from datasheet).

The gain is set to 11, resulting in a total system gain of 1100. Hence, the biosignal, including noise, cannot exceed 3mV otherwise clipping will occur on the 3.3V output rail. This is well suited for ECG, however for EEG the gain should be adjusted by increasing the ratio of the R3/R4 resistors. The signal conditioner has an optional integrated RLD which could be toggled to allow for 3 electrode mode, by using two jumper pins that would pull the *RLD* and *AC/DC* pins high or low, see Appendix A for the modified circuitry.

2.3.2 Analog-to-Digital Converter

Device	Unit Price on RS (£)	No of Channels	Resolution (bits)	Sampling Rate	Input-Referred Noise (uVpp)	Signal to Noise Ratio (dB)
ADS1299-x	46.66 (29.37)	8 (6)	24	16k	1	121
ADS1194CPAG	9.67	8	16	8k	12.2	97
ADS1298R	29.71	8	24	32k	4	112
MCP3464	4.97	8	16	153k	3.2	97

Table 3: Comparison of popular low-noise 16-bit delta-sigma analogue-to-digital converters.

The MCP3464 ADC was chosen, see datasheet, since it offers the highest sampling rate out of the ones evaluated and therefore can be over-sampled to improve signal quality. Additionally, it is of lowest cost with comparable signal to noise ratio and resolution to others.

Delta-sigma ADCs are designed to oversample far above the Nyquist frequency to improve signal quality. The oversampling ratio can be calculated as follows:

$$\text{Oversampling ratio} = (\text{Sampling rate of ADC}) / (2 \times \text{Maximum frequency of input signal}) \quad (4)$$

In the case of ECG, since the signal conditioner has a low-pass at 40Hz, therefore the Nyquist frequency would be as little as 80Hz. However, due to the nature of delta-sigma ADCs, the entirety of the 153kHz sampling frequency can be used by setting the OSR appropriately. I.e., when recording across eight channels at 1kHz, the OSR could be set to 16, and if only one channel was used, the OSR could be set to 76, thereby reducing the effects of noise and improving the accuracy of the digitized signal.

2.3.3 Electrodes

Typical single-lead ECG devices adhere to Einthoven's triangle which states that three electrodes must be placed on the body in a triangle, so that the vector generated by the heart, which behaves like a rotating dipole, can be correctly identified [82], resulting in different ECG morphologies. In typical single-lead ECG there is a third electrode, RLD, explained in 2.3.1. This setup follows the standard Lead I configuration, however the aim of the device is to eliminate the requirement for this third electrode to improve comfort. Therefore, a two electrode chest strap is used, relying on the CMRR of the amplifiers and filtering circuitry to attenuate noise.

Additionally, in clinical ECG, wet electrodes are usually used, such as silver-silver chloride, where a sticky conductive gel is required to adhere to the body. These types of electrodes have demonstrated to result in the highest signal quality [83]. In contrast, conventional dry electrodes use a single metal as a conductor between the skin and the electrodes, and do not need conductive gels. Because of the dry electrodes' easy setup, many types of ECG wearables such as wristbands, chest straps, and smart watches use a dry electrode-based ECG sensor, see Specification 1. The Polar H10's Polar Pro Strap uses a conductive electrode chest strap which has shown to outperform other wearables in signal quality [84] and will thus be used in addition to a 3.5mm connector which supports 3 electrodes, including the RLD.

2.3.4 Other Components

Microcontroller: The ESP32 was used, see datasheet, due to its high performance and low-cost. The device has two cores allowing for multi-threading, up to 3 SPI and 2 I2C buses for communication with the SD card, IMU and other peripherals, and has integrated WiFi and Bluetooth modules built-in. The device supports FreeRTOS which allows for easily extendable tasks to be run concurrently on either core and respond in a deterministic manner. See Section 2.4 for the breakdown of these tasks.

IMU: The IMU must be able to produce sufficient signal quality to detect different types of movements and thresholds, so that it can be compensated for in the stress classification model. Subsequently, the MPU6050 was chosen since it offers both 3-axis accelerometer and gyroscopic measurements, which has been used in numerous studies and is capable of distinguishing different types of gait [85]. The device is the most costly component, and after research, a lower-cost IMU could be used if precision of movement, such as gentle vs fast paced walking, does not need to be distinguished for accurate stress classification.

Power: The device uses a 820mAh, 3.7V battery with regulation and charging circuitry. Whilst in offline mode, the device uses 30mA, corresponding to 27.3hrs of battery life. During streaming over WiFi, the device uses 120mA (6.8hrs). The battery connector allows for replacement during long recordings.

2.3.5 PCB Design

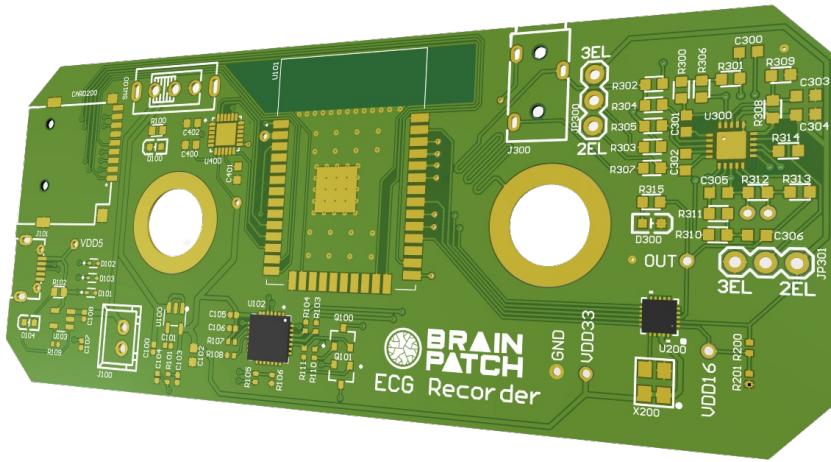


Figure 8: Finalised ECG Recorder Printed Circuit Board (PCB) - designed using Altium Designer. See Appendix A for the device circuitry and Table A.6 for the bill of materials.

The PCB design was aimed to be compact, whilst ensuring components were mounted on one side only to reduce assembly complexity and cost. Two holes were made through the PCB, where connectors were mounted to fit the Polar chest strap. An LED was placed in parallel to the signal output of the AD8232. This allowed for visual inspection of the heart beat and meant any poor connection could be easily identified.

PCB design best practices were followed, the most notable were:

- **Differential pair length tuning** was enforced to ensure that the electrode to signal conditioner trace distances were precisely matched to minimise distortion and improve common mode rejection.
- **Thermal vias** were used under the ESP32 to mitigate heat.
- **Power line track widths** were made wider (25 mil) to prevent thermal runaway.
- **Keep-out layer** was created near the ESP32's WiFi antenna to prevent interference.

2.4 Firmware

Wireless Communication

User Datagram Protocol (UDP) is a network communications protocol that is often utilised for time-sensitive applications and has been chosen over Transmission Control Protocol (TCP). Unlike TCP that ensures a connection is established and that the packages sent are successfully received through a checksum and acknowledgement, UDP offers a connectionless protocol for fast data transmission [86] that meets the requirements set in Section 2.1.

Architecture

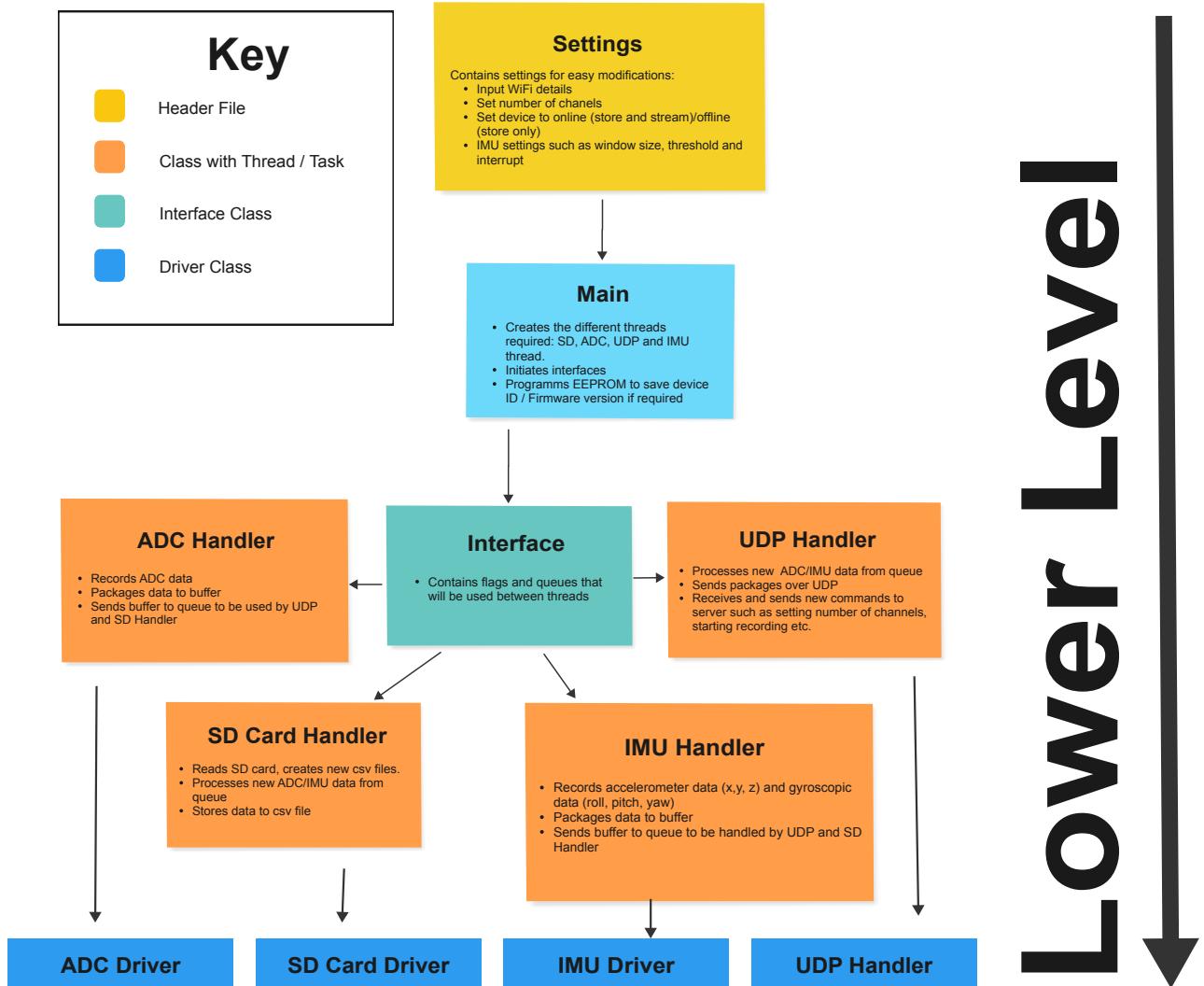


Figure 9: Firmware Architecture for the ESP32.

As discussed in Section 2.3.4, the ESP32 utilises FreeRTOS to arrange tasks or threads. The tasks for this device are for handling the ADC and IMU recording, SD card file storing and wireless communications and data transmission using UDP.

3 Experimental Method and Device Results

In this section, the experiment protocol, data acquisition, signal processing and classification implementation is discussed. Additionally, the results of the ECG recorder's design by means of signal quality will be outlined, since it is essential for signal processing and feature extraction.

3.1 Protocol and Data Acquisition

Due to the project's time constraints, it was unfeasible to conduct a controlled experiment, and thus, this preliminary study will serve only as a demonstration and to present initial findings. The experimental setup involved the participation of three male individuals aged between 21 to 22 years, with no known heart conditions or medication affecting the heart. All individuals signed the consent form, see Appendix A.



Figure 10: Participant wearing ECG recorder prototype with 3d-printed enclosure and Polar chest strap.

The participants wore the ECG recorder with the chest strap, and were instructed to wear the device while studying and record their stress level initially and whenever their stress changed during the recording. There was no induced stressor in this experiment and solely relied on the participants to naturally cycle through stressful states during their studies. Stress levels were recorded between 1 (extremely relaxed) to 10 (extremely stressed) which were subsequently mapped to low (0-4), medium (5-6), and high (7-10). The recordings with medium labels were removed for binary classification. In accordance to 2d of the design objectives, the participants were allowed to move, however the IMU was set to a threshold of 100mg which corresponds to the threshold of walking. When this threshold was reached, the timestamp would be recorded. To ensure that the participant's heart rate returned to a normal baseline level, the ECG was ignored for 5 minutes following the motion detection event. 20 recordings were obtained, lasting a total of 286 minutes.

Participant	Recordings	Duration (mins)	Stress Level Duration (%)		
			Low	Medium	High
1	1 to 9	80	0.0	82.1	17.9
2	10 to 13	76	34.0	62.2	3.8
3	13 to 20	130	15.6	62.2	22.2

Table 4: Stress level distribution per participant.

The stress classification performance from the ECG recorder device was cross-evaluated against the publicly available Spider-fear database [87] from the RCT conducted by Ihmig et al. The raw 100Hz

ECG signal from the study was captured using an open-sourced, low-budget device (BITalino) where three electrodes are placed according to standard lead II configuration. The EDA and RESP were also recorded using this device but were discarded in this study. The researchers kindly gave access to the Subjective Ratings file for the rated stress levels and the SB approach was followed whereby the last three minute of the biofeedback training phase was used for low stress, subjective ratings 1 and 2 for medium stress, and ratings 3 and 4 for high stress.

Database	Experimental Type	Participants	Configuration	Sampling Rate (Hz)	Stress Level (%)			Total Duration (mins)
					Low	Medium	High	
BrainPatch	Demonstrative	3	Bipolar Lead I Chest Strap	1000	19.4	61.1	19.5	286
Spider-Fear	RCT	53	Lead II w/ RLD	100	14.8	57.3	27.8	583

Table 5: Comparison of experimental design and stress level distribution for each database.

3.2 Methodology Overview

The stress classification experimentation was written in a Python Jupyter Notebook, see Appendix A. In high-level, the notebook first sorts through the raw ECG signal by segmenting and adding stress level labels. Secondly, key HRV, ECG and EDR features are extracted. The labelled features are then passed to the ML model for training and classification, and lastly evaluated. As discussed in Section 1.5, the CNN features are extracted automatically and do not require manual feature extraction.

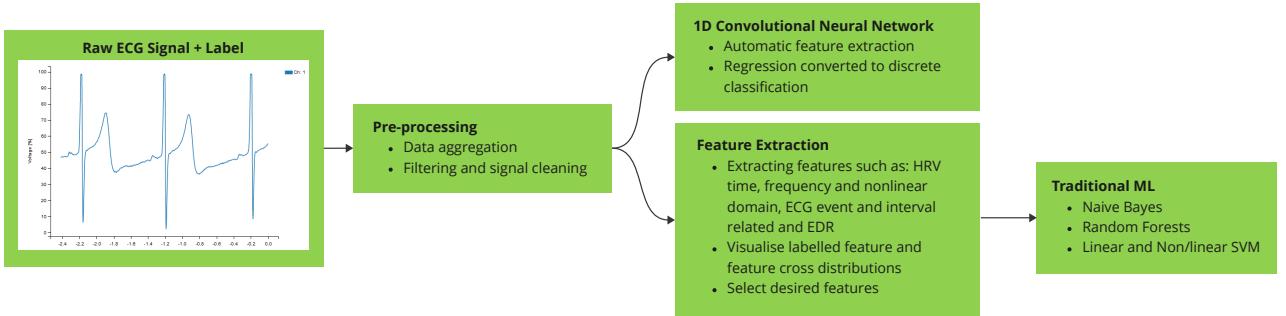


Figure 11: Overview of stress classification procedure using traditional and deep ML methods.

Apparatus and Resources Used: The libraries used for this experiment were: pandas, numpy, tensorflow, scikit-learn, imblearn, neurokit2, matplotlib and seaborn. More information is given in the notebook, see Appendix A. To minimise preprocessing and training time, the experimental analysis was conducted on a PC running an Intel 11th i9-11900 processor with 64GB of Ram and a 10GB NVIDIA GeForce RTX 3080 GPU with CUDA installed to use with TensorFlow.

3.3 Signal Analysis and Preprocessing

3.3.1 Signal Quality Analysis

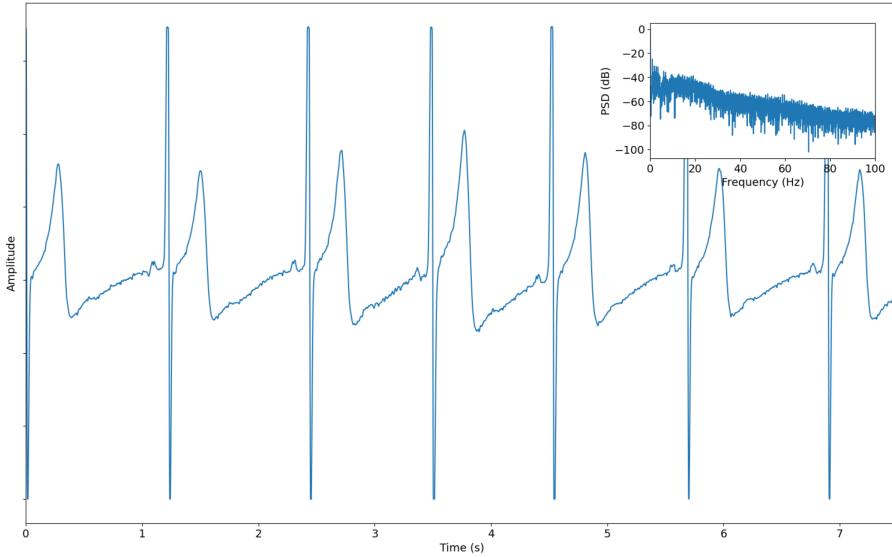


Figure 12: Raw ECG segment from the BrainPatch device during a participant recording.

The ECG waveform demonstrates that the device effectively suppresses noise, particularly from powerlines, resulting in clear visualization of the key waveforms. The gain is correctly set to maximise the ADC's resolution without capping the R peaks. However, there are several artifacts which do not adhere to the ideal ECG trace shown in Figure 2.

1. The P wave has an amplitude that is only slightly more than the surrounding noise.
2. The Q wave, which should have an amplitude of approximately 25% of the S wave, is smaller and less pronounced.
3. The signal shows to be isoelectric where the S wave has an amplitude comparable to that of the R wave.
4. Between the T and the next P wave, the signal increases in amplitude, showing device saturation.

In comparison, the signal quality of the Spider-fear devices, see Figure A.8, shows to have more pronounced P and Q waves. It is unclear if the wave following the T wave, is an inverted U wave or a continuation of the T wave. From the Power Spectrum Density (PSD) subplot, it appears that the device used a high order 50Hz low-pass filter to attenuate any signal exceeding this. The R peaks are sloped due to the low sampling rate, in contrast to BrainPatch's signal which shows a more rounded curve. This attribute will be beneficial for precise R peak detection and amplitude measurement, resulting in a more precise HRV, and EDR signal respectively.

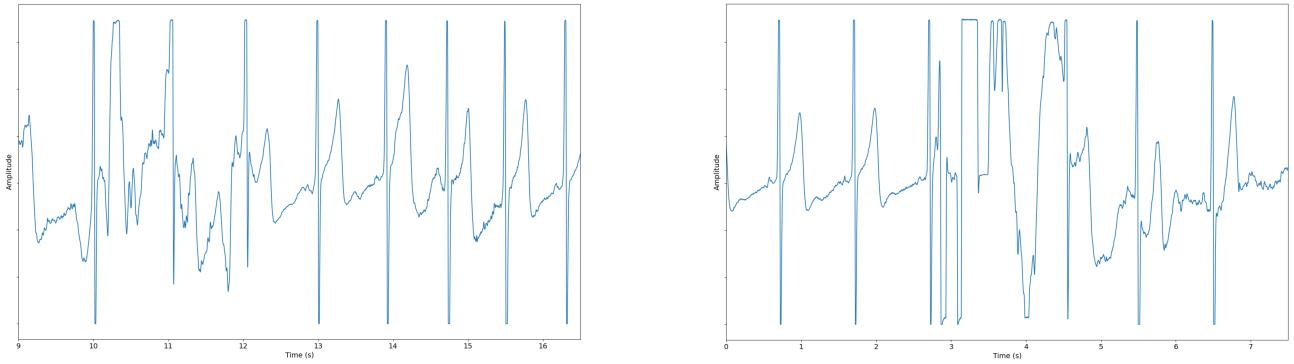


Figure 13: Segments of recordings 3 (a) and 14 (b) that contain large artifacts from noise or saturation.

To ensure the quality of the ECG signals, each recording was carefully inspected for any noticeable artifacts that might have resulted from poor electrode contact. Unfortunately, during BrainPatch recordings 3, 14, 15, and 18, the ECG signals became noisy or saturated and were therefore excluded from the analysis. Additionally, due to poor signal quality, the ECG signal of one participant (VP70) from the Spider-fear study was also removed. The ECG signals were then automatically labelled using the Events file and Subjective Ratings file for BrainPatch and Spider-fear databases respectively.

The ECG recordings from the BrainPatch device have corresponding timestamps which were interpolated to ensure a 1kHz sampling rate with no jitter was achieved. As stated in the paper, the Spider-Fear data were not able to be interpolated since the timestamps were corrupted. Interpolation was also used to downsample the ECG to find the minimum sampling rate required before a lowering in classification accuracy, see Objective 3c.

3.3.2 Filtering

In addition to the signal conditioner filtering, see Section 2.3.1, digital filtering was also required to remove the impact of artifacts thereby improving feature extraction reliability. The frequency of an ECG signal and its artifacts are shown below.

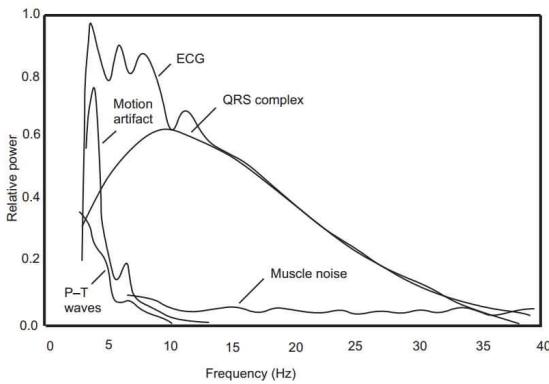


Figure 12.1 Relative power spectra of QRS complex, P and T waves, muscle noise and motion artifacts based on an average of 150 beats.

Figure 14: The relative power spectrum of an ECG signal and its artifacts [64].

Neurokit2 [88] is an advanced biosignal processing toolkit, containing several methods for cleaning and analysing ECG. The *ecg_clean* function was used for filtering which implements a 5th order 0.5Hz high-pass butterworth filter to reduce baseline wander and motion artifacts, a 50Hz notch powerline filter to eliminate mains induced artifacts and a 150Hz low-pass filter for high frequency artifacts such as muscle.

Pan Tompkins is a widely used algorithm that implements a 5-12Hz pass band for the cleaning and feature extraction ECG [89], however has some limitations. Instead, a 5th order 0.5Hz high-pass

butterworth filter was implemented which provided two added benefits. Firstly, low-frequency components of the ECG below 5Hz are important and should not be attenuated, otherwise features may be lost. A 40bpm heart rate for example, would result in an R-R frequency of 0.67Hz and it is important that the amplitude of the peaks are not reduced from filtering to ensure detection and amplitude can be detected for HRV and EDR respectively. Secondly, the butterworth filter provides a more linear phase and flat response within each bandwidth, compared to other low-pass filters, therefore waveform phase and frequency distortion will be prevented.

An additional high pass filter was used to remove high frequency components such as muscle artifact and external interference. Many ECG analysis studies use 50Hz as the stop band, however this prevents information from higher frequency features of the ECG, such as the QRS complex, to be extracted. Stated by the American Heart Association "to measure durations and amplitudes accurately in adults, an upper-frequency cutoff of at least 150 Hz is required" [90]. Therefore to prevent any loss of information, particularly for the raw signal which will be inputted into the CNN model, the low-pass filter was set to 150Hz. Since there are no frequencies above 50Hz for the Spider-Fear ECG, these features are already lost.

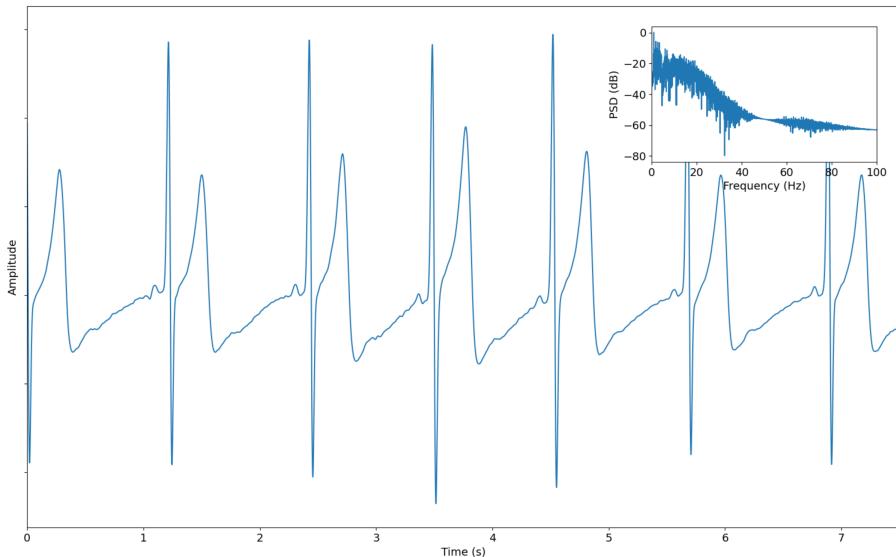


Figure 15: Cleaned BrainPatch ECG segment after filtering.

The PSD subplot shows that the 50Hz band has been severely attenuated, reducing the noise from the powerline. In addition, the DC components, motion artifacts and baseline drift artifacts below 0.5Hz has been removed. This results in frequencies between 0.5-20Hz having a higher relative power, which should improve feature detection, thus improving accuracy for traditional ML models, in addition to reducing the overall noise which may be learnt by the CNN, resulting in poor generalizability.

3.3.3 Data Preprocessing

Data Augmentation: Similar to Tzevelekakis et Al., both datasets are relatively small and are insufficient for training a deep neural network. Therefore, a sliding window was applied to the cleaned ECG data, where each segment had a set window length and overlap. This was written manually, however retrospectively, Neurokit's *epochs_create* function could have been used. Overlapping is a type of synthetic oversampling whereby more data can be artificially generated which has been shown to improve accuracy, particularly in deep learning models [91]. However, there is a risk of data leakage caused by this and care must be taken when splitting the data to minimise this. Additionally, to ensure consistency when validating, the traditional ML methods will also be trained on this augmented data. As stated in Objective 3b, the window length will aim to be minimised for reduced computational time.

Class Imbalance Correction: The class imbalance for both datasets are significant, see Table 5, and must be addressed to ensure the minority classes (Low and High) are equally represented. Synthetic Minority Oversampling Technique (SMOTE) was chosen, whereby new minority class samples are created along the line segments that connect existing minority class samples [92]. By oversampling the data, this ensures that there is sufficient data to train the models, particularly for the CNN which demands a high number of samples.

Normalisation and Imputation: Over-fitting often occurs if the data is not scaled correctly and can also prevent convergence for neural networks [92] or lead to poor accuracy for SVM [51]. To combat this, Scikit-learn’s *regularisation* function was used whereby the values of a vector are scaled so that the sum of the squares of the elements is equal to 1, resulting in a float between 0 and 1. For values that were either missing, NaNs or infinities, imputation was used that took the mean of the feature column, using Scikit-learn’s *SimpleImputer*.

Train-test Split Procedure: To minimise data leakage, the training data was first randomly split into two datasets - train-test and validation using a ratio of 80:20. The sliding window was then applied to the train-test dataset only and then split again using Scikit-learn’s *StratifiedShuffleSplit* function whereby the data is randomly splitted into folds, with a ratio of 70:30, containing a representative sample of each class for both train and test. This ensured that no validation data was used during training.

3.4 Feature Extraction

The extracted features can be classified into three types, namely HRV, ECG, and EDR. For each type, the method used for feature extraction will be described, the reliability of feature extraction of the signal, acquired by the device will be evaluated.

HRV features were obtained using Neurokit’s *ecg_intervalrelated* function, which follows the HRV analysis techniques detailed in Pham et al. [93]. A more concise list of features were obtained from Neurokit’s *hrv_time* and *hrv_frequency* functions. From this function, 20, 9, and 40 features for time, frequency and nonlinear domains respectively were extracted. From these features, the most significant features from Castaldo et al. were selected:

Domain	Feature	Description
Time	Mean NN	The average of the RR intervals between successive QRS complexes, excluding anomalies
Time	Max NN	The maximum of the NN intervals
Time	Min NN	The minimum of the NN intervals
Time	SDRR	The standard deviation of the RR intervals in the ECG signal
Time	RMSSD	The root mean square of the successive differences between RR intervals in the ECG signal
Time	pNN50	The percentage of successive RR intervals that differ by more than 50 ms in the ECG signal
Frequency	LF	Total spectral power of the low-frequency band (0.04-0.15 Hz) in the HRV signal
Frequency	HF	Total spectral power of the high-frequency band (0.15-0.4 Hz) in the HRV signal
Frequency	LF/HF	The ratio of the low-frequency power to the high-frequency power in the HRV signal
Nonlinear	DFA α_1	The scaling exponent of the detrended fluctuation analysis (DFA) curve in the HRV signal
Nonlinear	D2	A measure of the correlation dimension of the HRV signal
Nonlinear	SD1	The standard deviation of the beat-to-beat variability of the Poincare plot in the HRV signal
Nonlinear	SD2	The standard deviation of the long-term variability of the Poincare plot in the HRV signal
Nonlinear	SampEn	A measure of the complexity of the HRV signal
Nonlinear	ShEn	A measure of the Shannon entropy of the HRV signal
Nonlinear	REC	A measure of the recurrence of patterns in the HRV signal
Nonlinear	lmean	The mean length of the diagonal lines in the recurrence plot of the HRV signal
Nonlinear	lmax	The maximum length of the diagonal lines in the recurrence plot of the HRV signal

Table 6: The most significant features for stress classification ($p < 0.05$), as identified by Castaldo et al., extracted from Neurokit’s HRV features.

All the HRV features rely only on R-R intervals, which were validated to ensure the intervals were in the range of a healthy participant 30-200 bpm. No intervals showed to be outside of this range and

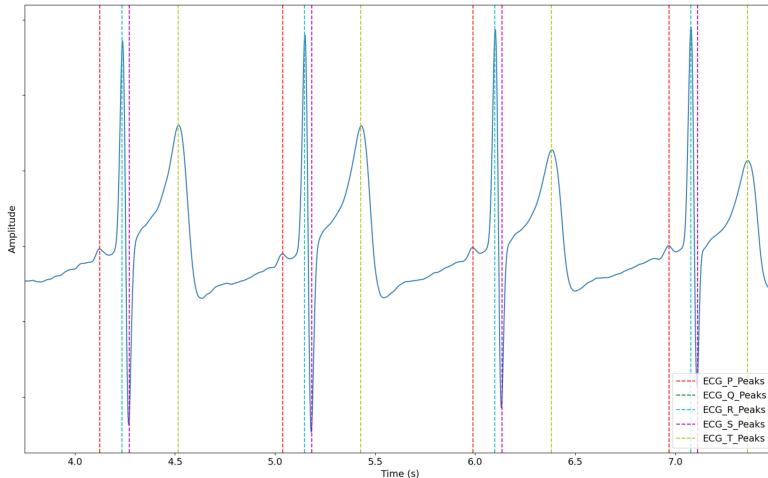
thus HRV features show to be extremely reliable for both devices.

ECG features were extracted using three functions:

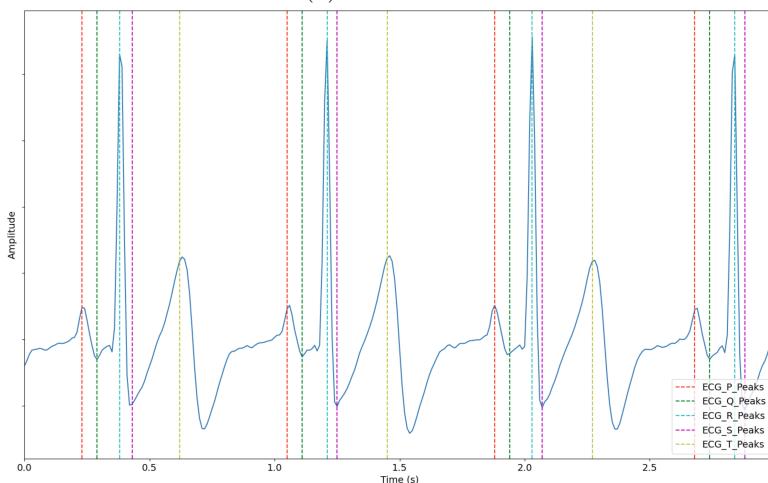
1. *Neurokit's ecg-process*: Extracts key locations of the waveform such as the onset, offset and peak for the PQRST waves, in addition to the atrial (PQ) and ventricular (QT) phase.
2. *Wave_analysis*: Using the locations calculated from *ecg_process*, calculates the interval and duration of the features. This function accounts for the first index to be either the onset or the offset, when determining its duration. Before calculating the mean and standard deviations, the durations and intervals are checked against a healthy range, i.e. the maximum duration for a P-wave is 120ms [94].
3. *Calc_PSD*: The Fourier Transform is calculated, and subsequently the logarithm of summed total power in 10Hz bands, from 0 to the nyquist frequency, is calculated to produce a binned frequency power spectral density of the ECG segment.

Type	Feature	Description
Time	ECG_P_Peaks_Interval Mean/SD	The mean / standard deviation of the interval between successive P peaks
Time	ECG_Q_Peaks_Interval Mean/SD	The mean / standard deviation of the interval between successive Q peaks
Time	ECG_S_Peaks_Interval Mean/SD	The mean / standard deviation of the interval between successive S peaks
Time	ECG_T_Peaks_Interval Mean/SD	The mean / standard deviation of the interval between successive T peaks
Time	ECG_P_Duration Mean/SD	The mean / standard deviation of the duration of P waves
Time	ECG_T_Duration Mean/SD	The mean / standard deviation of the duration of T waves
Time	ECG_QT_RMSSD	The root mean square of the successive differences of QT wave durations (VRV feature)
Time	ECG_Rate_Trend_Linear	The parameter corresponding to the linear trend of the ECG
Frequency	ECG_PSD	The Power Spectrum Density logarithm of summed total power in 10Hz bands
Nonlinear	ECG_Rate_Trend_Quadratic	The parameter corresponding to the curvature of the ECG
Nonlinear	ECG_Rate_Trend_R2	The quality of the quadratic model used to fit the ECG trend

Table 7: Extracted ECG features such as atrial and ventricular phase, and wave interval and duration.



(a) BrainPatch



(b) Spider-fear

Figure 16: ECG segment with peaks extracted using Neurokit’s `ecg_process` function for both devices.

The ECG signals from both devices demonstrate that the waveform peaks can be extracted, however to different extents. The BrainPatch device shows to consistently derive the PQR and T peaks, which allows for measurement of the P, PR and T intervals. However the Q peak could not be reliably detected, using Neurokit’s, Pan Tompkins, or Elgendi et al. [95] methods. Unfortunately, this limitation means that VRV features cannot be obtained due to the inability for QT interval measurements. The Spider-fear device shows to accurately identify all peaks, however precision is lacking, due to the device’s low sampling frequency. The sloped R peaks cause the maxima to be located either side of the peak, this results in a variation of approximately 5ms, however it is unlikely to impact HRV metrics too considerably.

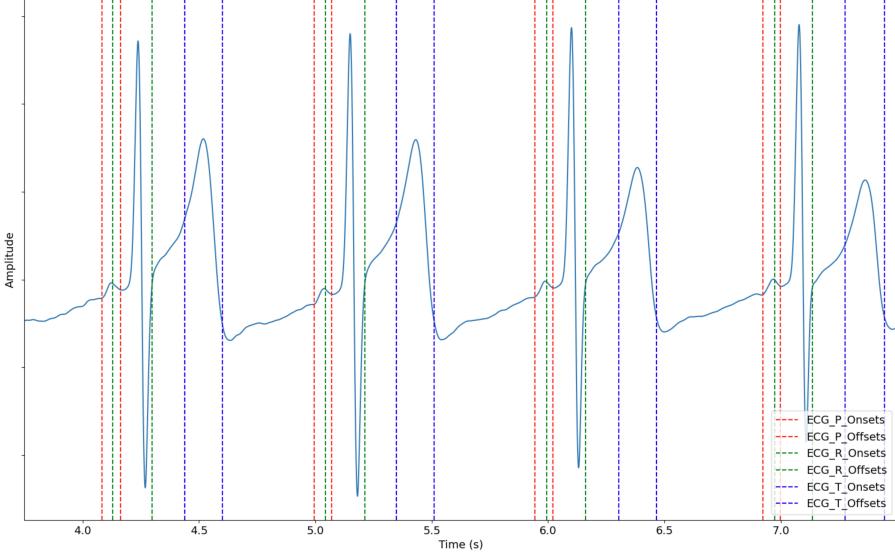


Figure 17: ECG segment with onset and offset waves extracted using Neurokit’s *ecg_process* function.

The BrainPatch device shows to reliably identify the onset and offset of the P, R and T waves. This is used to calculate the duration of the P and T waves. The R wave duration is already calculated in the HRV features. The *ecg_process* function is not designed to identify the S wave, however this duration is not required, as little evidence has been found to show that stress impacts the S wave.

Feature	Duration (ms)	
	Healthy Normal	Average Recorded
P wave duration	<120	87
PR interval	120-200	135
QRS duration	60-100	Unavailable
QTc interval	350-450	Unavailable
T wave duration	160	130
ST segment	80-120	93

Table 8: BrainPatch ECG average feature durations using the peaks, onsets and offsets generated, compared against healthy durations [96, 97, 98].

To verify the reliability of the peak, onset and offsets extracted, the metrics of key features were measured to evaluate if they fell inside a healthy range. All metrics that did not rely on the Q wave fell inside this range and thus features using P,R, and T waveforms can be relied on.

EDR features were extracted using Neurokit’s *ecgrsp* function, which implements the HeartPy algorithm [99] to accurately identify the R peak amplitudes, and subsequently derive the RESP signal, like in Figure 3. The local minima and maxima of the RESP signal was then found to identify the mean duration of breaths.

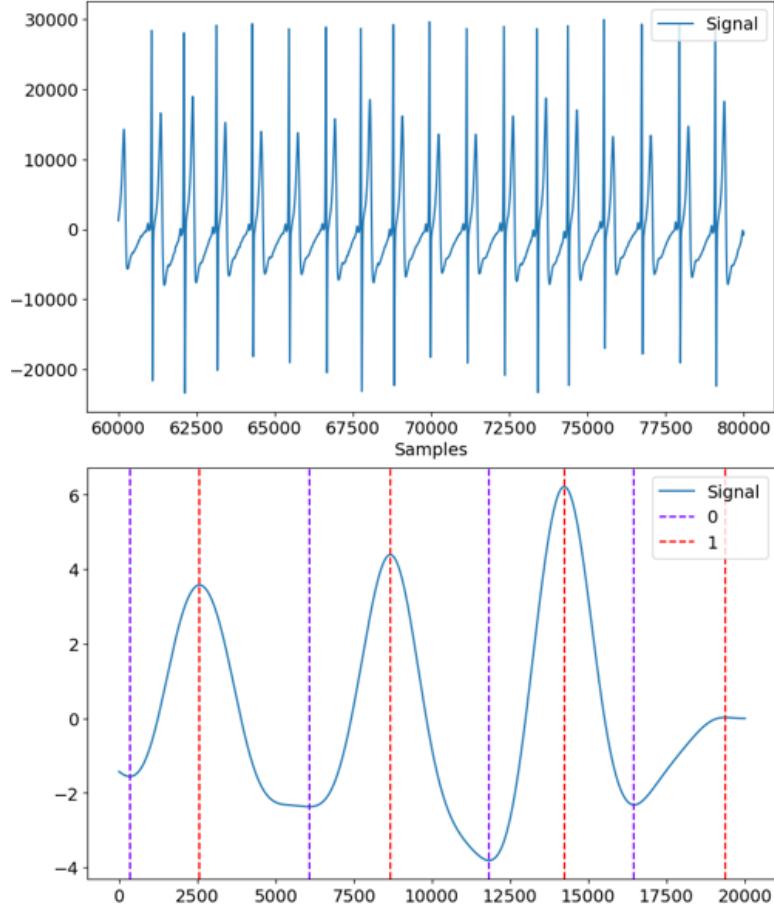


Figure 18: 20s ECG segment (top) with its corresponding RESP signal (bottom) using EDR. '0' indicates a local minimum and '1' a local maximum.

The maxima of the RESP signal shows to correlate with the maximum peak R amplitude, and the interval between each maxima can be used as an EDR feature, mirroring that of Figure 3. However, in this segment, the local minima shows to vary considerably and does not yield a reliable interval, and has thus been disregarded as a feature.

4 Classifier Implementation and Results

This section details how each model was implemented and tuned using the BrainPatch dataset. The models will then be compared with each other to evaluate their accuracy, generalizability, and computational performance. Additionally, the window length and feature selection will be discussed as this will dictate computational performance.

For each model, a 5-fold cross-validation of the stratified splits were used to evaluated its performance. The following metrics were used:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP and TN are true positive and true negative classifications and FP and FN are false positives and negatives respectively, which will be displayed in confusion matrices.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (6)$$

Additionally an F1-score will be used, which incorporates two important metrics: precision (true positive predictions among all positive predictions) and recall (proportion of true positive predictions among all actual positives).

4.1 Naive Bayes

The Naive Bayes model was first tuned using Scikit-learn's *GridSearchCV* function, whereby the model can be trained on each combination of the hyperparameters and returns the best performing set of hyperparameters. The type of classifier used, which are variants of the Naive Bayes algorithm that assume different probability distributions for the input features. The alpha hyperparameter, or additive smoothing, prevents a zero probability from being predicted if the feature is not available, with a higher value increasing the smoothing [100]. Additionally, as described in Section 1.5, the probabilities of the training set's classes can be given to use in Bayes' theorem to compute the posterior probabilities for a Gaussian distribution.

Classifier	Parameter Grid
Multinomial	{'alpha': [0.01, 0.1, 0.5, 1.0, 5.0, 10.0]}
Bernoulli	{'alpha': [0.01, 0.1, 0.5, 1.0, 5.0, 10.0], 'binarize': [0.0, 0.5, 1.0]}
Complement	{'alpha': [0.01, 0.1, 0.5, 1.0, 5.0, 10.0]}
Gaussian	{'priors': Gaussian_priors}

Table 9: Grid search parameters for each Naive Bayes classifier type.

From the grid search, the best parameters were chosen for each classifier using the highest achieving accuracy, and were then re-evaluated with a 5-fold cross validation to obtain a more reliable measurement.

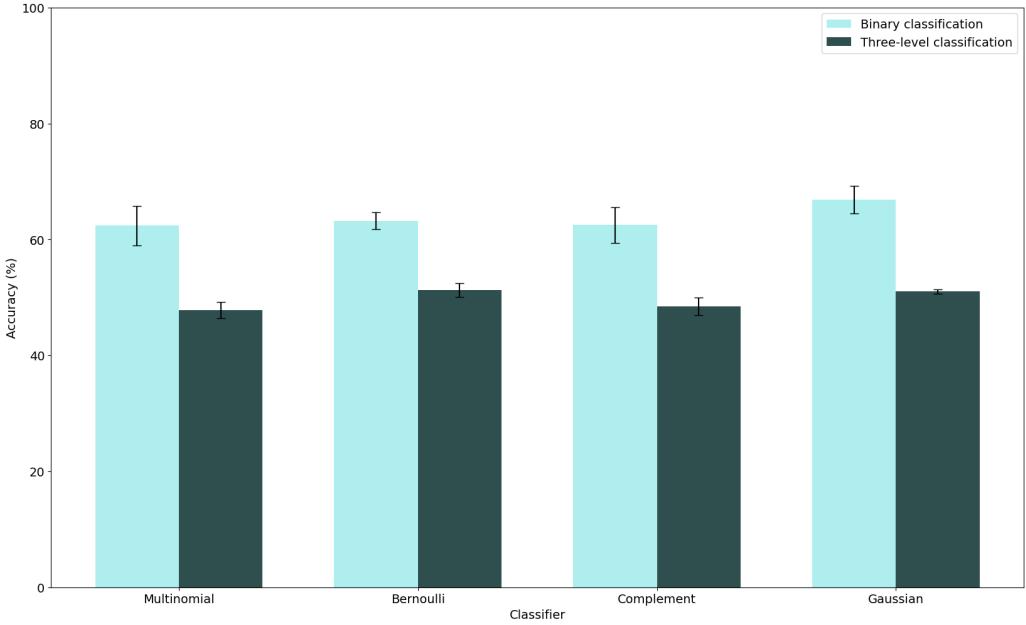


Figure 19: Comparison of optimised classifiers for Naive Bayes.

The Multinomial and Bernoulli classifiers show to be biased to predict one class over another, leading to a low F1 score and is not suited for classification. The Gaussian classifier demonstrates to be most accurate, particularly for 3 level classification, and carries the smallest deviation compared to Complement. Since a Gaussian classifier assumes all features are normally distributed, the model was re-classified using logarithmic scaled features, however this yielded a lower accuracy and demonstrates feature distributions are closer to normal distribution than log-normal distribution.

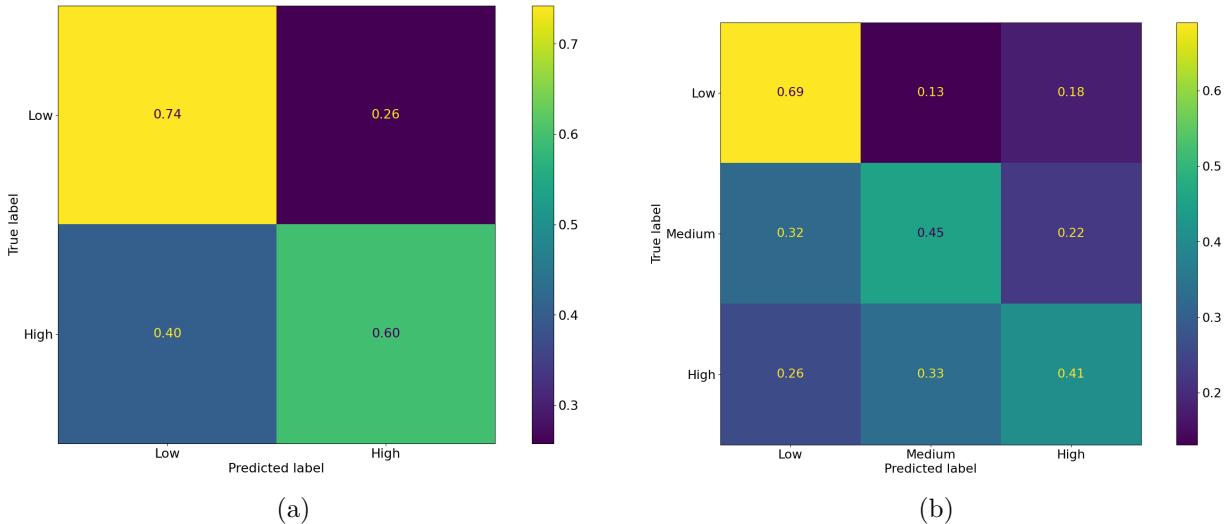


Figure 20: Confusion matrices for binary and three-level classification for Naive Bayes.

Both classification types demonstrate that low-stress is more easily separable, however Naive Bayes struggles with the other stress levels.

4.2 Support Vector Machine

Similar to Naive Bayes, the *GridSearchCV* function was used to determine the best hyperparameters for each SVM kernel.

Kernel	Parameter Grid
Linear	{'C': [0.1, 1, 10, 100, 1000]}
Polynomial	{'C': [0.1, 1, 10, 100, 1000], 'degree': [2, 3, 4], 'gamma': [0.1, 0.01]}
Radial basis function	{'C': [0.1, 1, 10, 100, 1000], 'gamma': [0.1, 0.01, 0.001]}
Sigmoid	{'C': [0.1, 1, 10, 100, 1000], 'gamma': [0.1, 0.01, 0.001], 'coef0': [0.1, 0.01, 0.001]}

Table 10: Grid search parameters for each SVM kernel.

Where the hyperparameter C is inversely proportional to the L2 regularization rate, which affects how generalizable the model is to new data. *Degree* controls the order of the polynomial function used to generate the decision boundary, and *gamma* controls the smoothness of this boundary, thereby also affecting generalizability. The tuned hyperparameters were ' C' : 1000, ' $degree'$: 4, ' $gamma'$: 0.1 and ' C' : 1000, ' $gamma'$: 0.1 for Polynomial and RBG respectively.

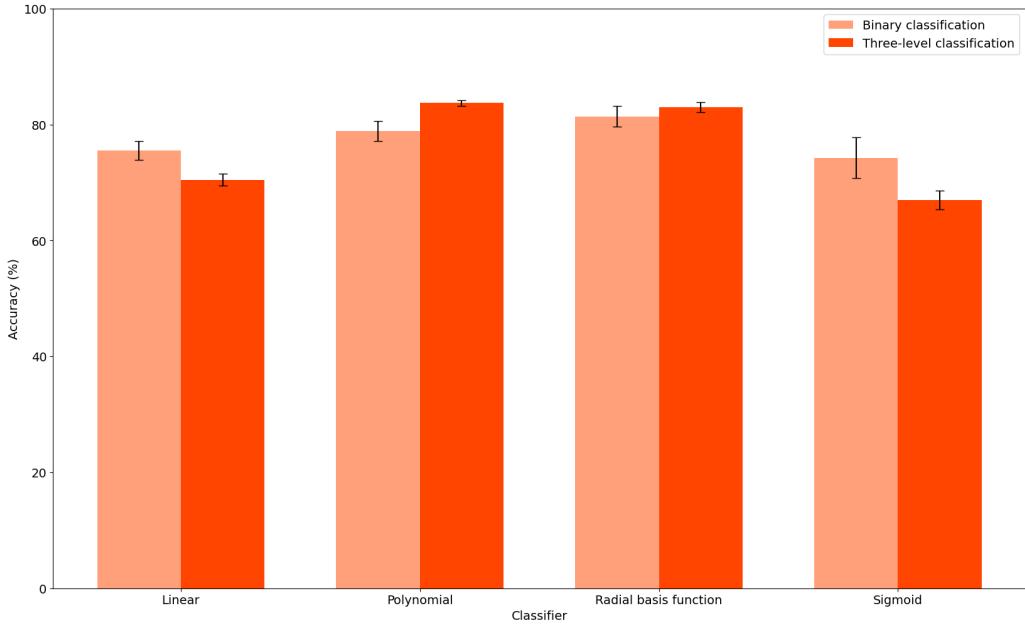


Figure 21: Comparison of optimised kernels for SVMs.

As expected, the linear classifier did not perform. From the nonlinear kernels, the sigmoid kernel performed less accurately due to its more simplistic decision boundary, and with higher deviation, likely due to overfitting in the test data. The polynomial and RBF kernels performed comparatively with RBF achieving 2% higher binary classification and polynomial achieving 2% higher three-level classification, with a reduction in deviation in both. Interestingly, and in contrast to Naive Bayes, three-level classification performed above the binary classification for these two kernels.

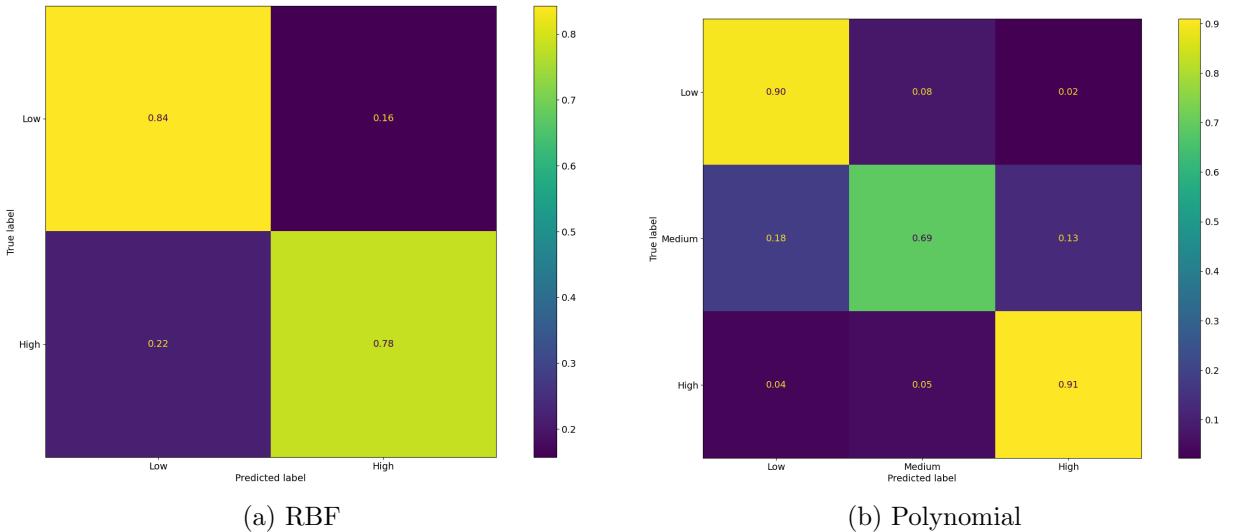


Figure 22: Confusion matrices for binary and three-level classification using SVM.

Both classification types demonstrate that SVM can differentiate the stress levels to a reliable extent, particularly for low and high stress, however medium stress only achieved 69% accuracy. There is a smaller percentage of false negatives for low and high stress using the three-level classifier. SVM demonstrates significant improvement to Naive Bayes, both in accuracy and deviation.

4.3 Random Forest

Criterion	Parameter Grid
Gini Index	{'n_estimators': [10, 50, 100, 250, 500], 'max_depth': [None, 2, 3, 4, 5, 7], 'min_samples_split': [2, 5, 10], 'max_features': [1, 'sqrt', 'log2', None], 'min_samples_leaf': [1, 2, 4], 'bootstrap': [True, False]}
Entropy	
Logarithmic Loss	

Table 11: Grid search parameters for each Random Forest criterion.

The hyperparameters were calculated for each criterion, which dictates how impurity is measured until the trees reach some stopping criteria [51], in this case if the impurity is less than `min_samples_split`. `Max_depth` sets the depth of the tree, unless the tree purity is lower than `min_samples_split`. `Max_features` sets the number of features to consider when looking for the best split, which affects the balance between classifying nonlinear features and overfitting. The other features are less impactful to performance but should still be tuned.

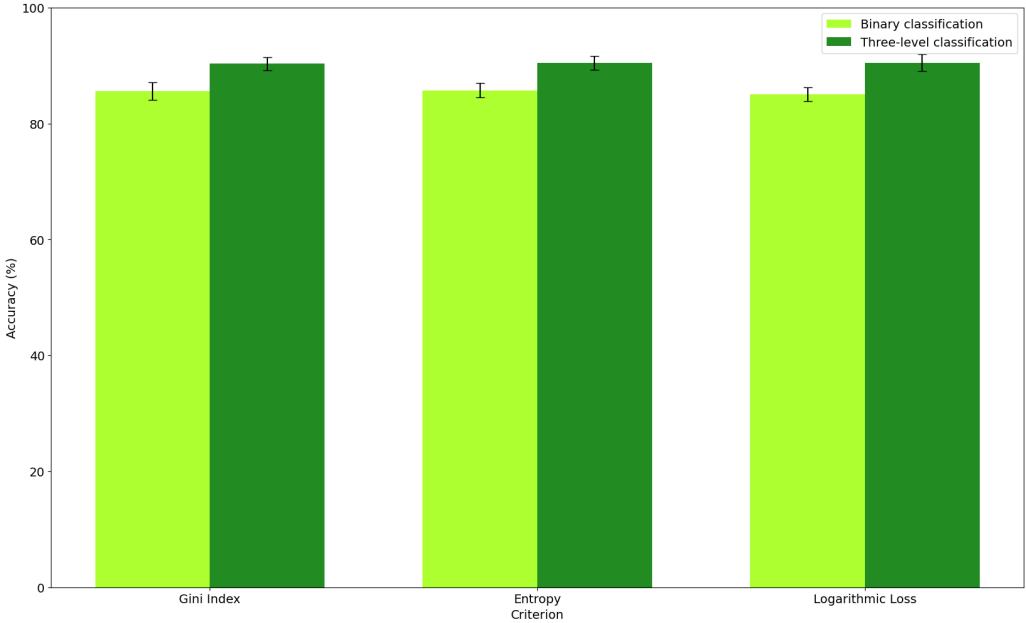


Figure 23: Comparison of optimised criterion for Random Forest.

The criterion shows to have little effect on performance, however entropy was chosen since it has the least deviation and therefore is less to be more generalizable.

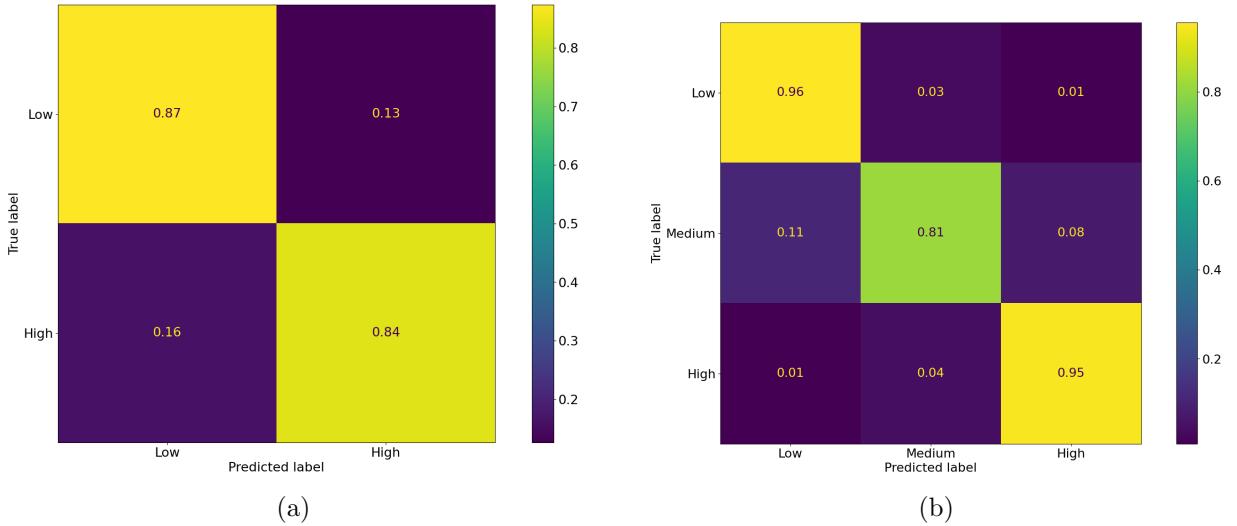


Figure 24: Confusion matrices for binary and three-level classification using Random Forest.

Similar to SVM, three-level classification outputted fewer false negatives for low and high stress in the three-level classifier. Random Forest shows to outperform SVM, particularly for medium classification accuracy, lowering false negatives in this class significantly.

4.4 1D Convolutional Neural Network

The implementation of the 1D CNN follows that of Tzevelekakis et Al. The input layer is an ultra-short (3s) raw ECG segment, which is convoluted by a filter of size, that on average, spans the PQRST waves (0.6s). For example, an 0.6s kernel size corresponds to 600 samples for the 1kHz BrainPatch signal. Crucially, the max-pooling layer, is of a size comparable to one ECG period (0.8s), thereby guaranteeing that the model can extract the peaks regardless of the peak point position, preventing the peak phase difference problem, whereby the model will not learn without changes to the phase of the peaks. A flattening layer converts the output to a 1D array of nodes; a dropout regularization

layer reduces overfitting by randomly dropping neurons every epoch. Two fully connected layers with activation functions *relu* and *softmax* were used to converge the neurons to the final output of either 2 or 3 classes for binary or three-level classification respectively. This totaled X parameters for a 1kHz configuration.

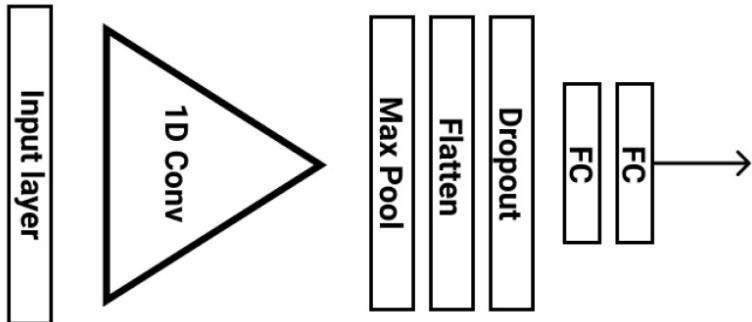


Figure 25: Single 1D-CNN Architecture from the Tzevelekakis et Al. [47].

TensorFlow's Keras library was used to create this architecture, train on the train and test set and predict on the validation set. The learning rate, epochs, batch-size and dropout rate were tuned for each window size, with the aid of a learning curve that illustrated the learning rate and amount of overfitting of the model.

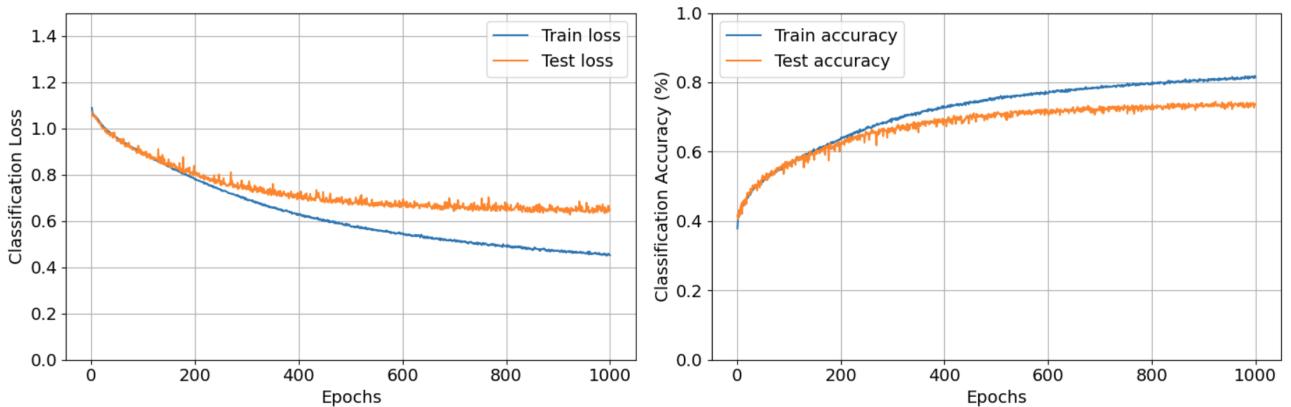


Figure 26: Learning curve for a 3s segment after 1000 epochs.

During training, the learning curve was extremely fluctuated, however a reduction in the learning rate and batch-size, and increase in epochs improved this. Early stopping was performed at just over 1000 epochs since the test accuracy/loss shows little improvement, and additional training would yield to considerable overfitting. Increasing the dropout rate further did not improve the test accuracy since the model would stop learning the model at all.

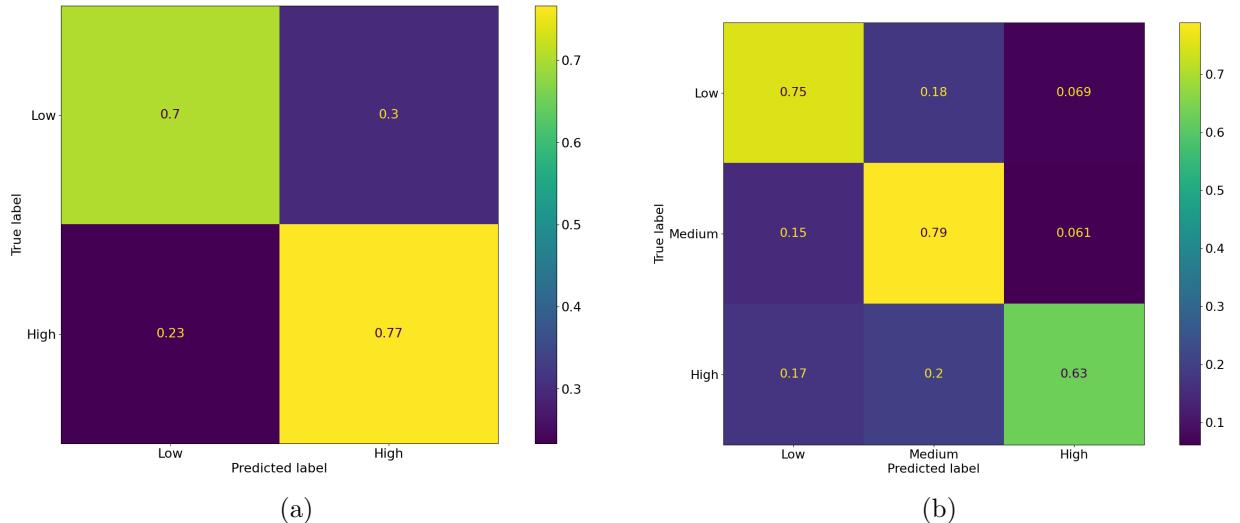


Figure 27: Confusion matrices for binary and three-level classification for 1DCNN.

The 3 second segment achieved an accuracy of 72.3% and loss of 0.48 for three-level classification, however struggled on the high stress classification. The optimal performance was observed at medium stress levels, which is likely attributed to the availability of a significantly higher number of samples. This could suggest that the SMOTE correction did not yield sufficient additional information, however low stress accuracy was still achieved in spite of this. Conversely, binary classification performs better on the high stress features, perhaps due to the model’s non-deterministic nature caused by the dropout layer and achieved 73.5% overall. Generating additional synthetic features by increasing the window overlap above 10%, improved the learning of the training set, however resulted in severe overfitting and did not improve test accuracy, see Figure A.12.

4.5 Model Comparison

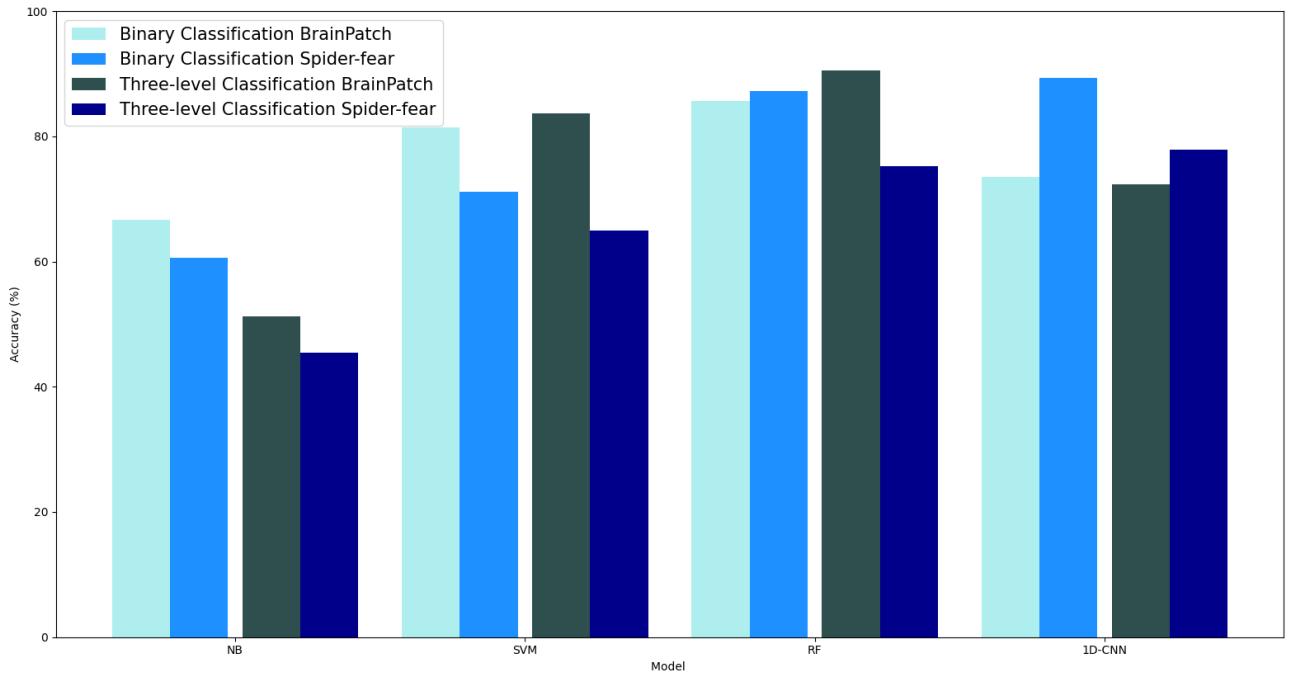


Figure 28: Model performance comparison for binary and three-level classification between the two datasets using a 30s window for traditional models and a 3s window for the CNN.

The model results for BrainPatch are taken from their corresponding sections above and the same method was implemented to tune each model for the Spider-fear dataset. See table A.1 for the complete accuracy and F1-Scores. Overall, the BrainPatch dataset was easier to classify over the spider-fear dataset for traditional methods, particularly for the SVM model which show a significant reduction in accuracy. Conversely, higher accuracies were obtained using the spider-fear dataset on the 1D CNN. This could be attributed to the dataset’s longer recording durations, which are particularly advantageous for neural network training. Nonetheless, Random Forest resulted in the highest score for both classification types and datasets, achieving 85.7% and 90.3% for binary and three-level respectively using the BrainPatch dataset.

Model	Runtime Normalized to Random Forest Three-Level Classification	
	Binary Classification	Three-level Classification
NB	0.139	0.160
SVM	0.146	0.962
RF	0.585	1.000
1D-CNN	11.210	11.210

Table 12: Normalized runtime comparison for each model.

The runtimes for each model show to vary dramatically, not only between models but also with classification type. For example, SVM binary classification is over 6 times faster than three-level classification, and comparable to its Naive Bayes equivalent. Random Forest shows to be similar in runtime to SVM for three level classification. The 1D CNN has a far higher runtime than traditional models, however this is due to the high 1kHz sampling rate which results in a larger number of parameters. See Section 4.7 for how runtime improves with sampling rate reduction.

4.6 Feature Selection

Using all the features listed in Section 3.4 for the traditional ML models is unrealistic since it requires significant computation. Thus, features must be carefully selected that are most separable between the different stress levels. One such method is to use a Kernel Density Estimate (KDE) plot that helps visualise the continuous probabilities of features for each class.

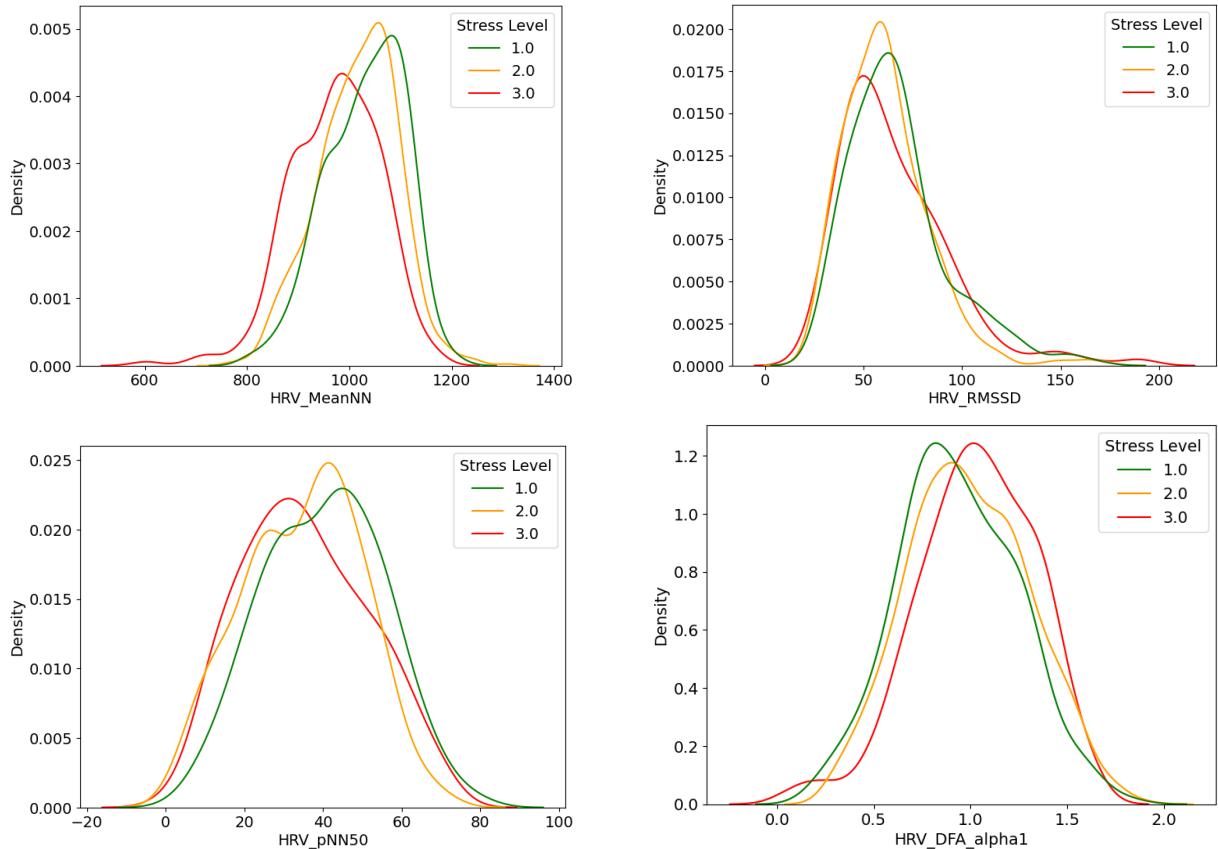


Figure 29: The most visually separable HRV features.

The mean, maximum and minimum NN, demonstrate to have large separability and behaves as expected, with shorter NN intervals corresponding to higher stress. Similarly, pNN50 demonstrates high correlation with stress whilst pNN20 has significantly lower. DFA also shows to be an excellent feature, particularly for binary stress classification.

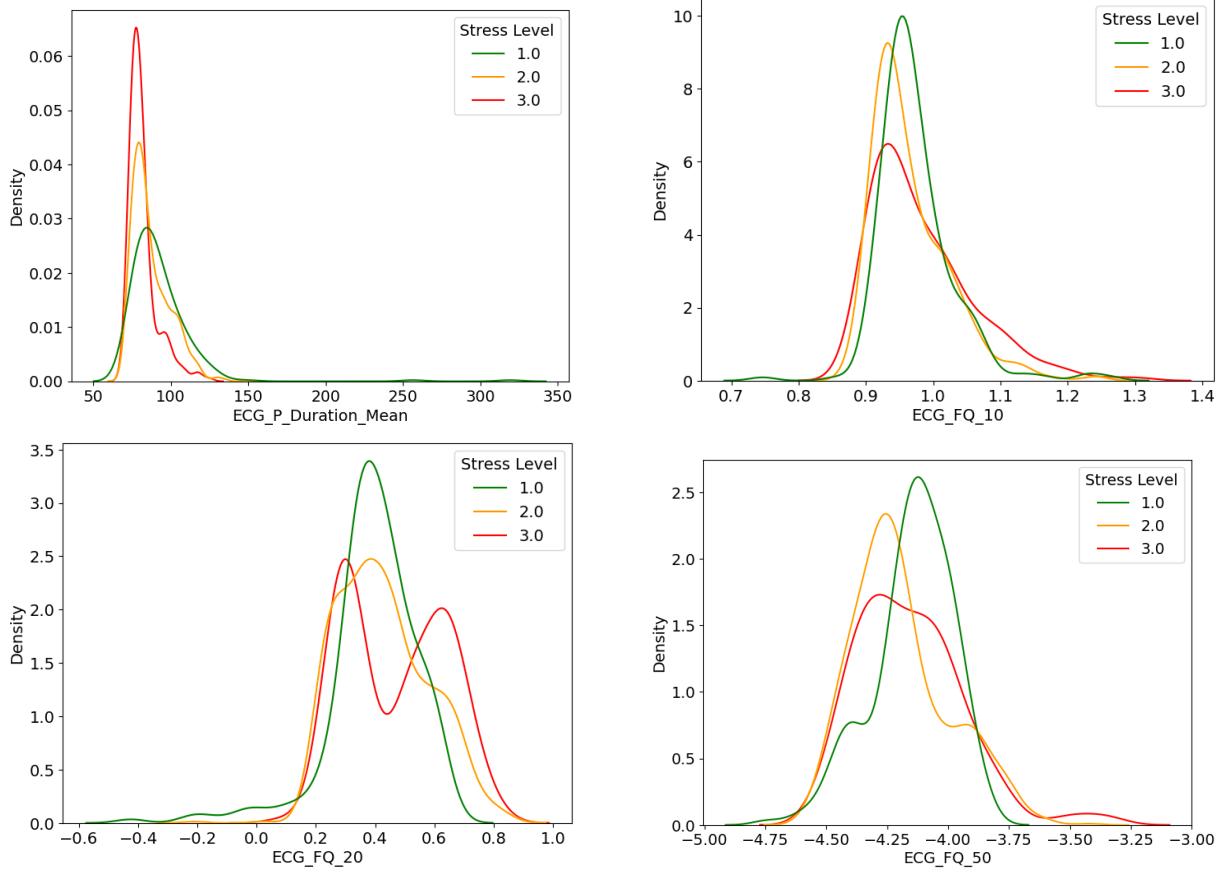


Figure 30: The most visually separable ECG features.

Out of all the intervals and durations of the ECG features, the P wave duration exhibits the greatest degree of separability. The frequencies 10, 20 and 50Hz also show to vary significantly between classes, however it is unclear if this is due to stress or artifacts from the recording since the Spider-fear dataset do not follow the same curvature or relationships between classes, see Figure A.10.

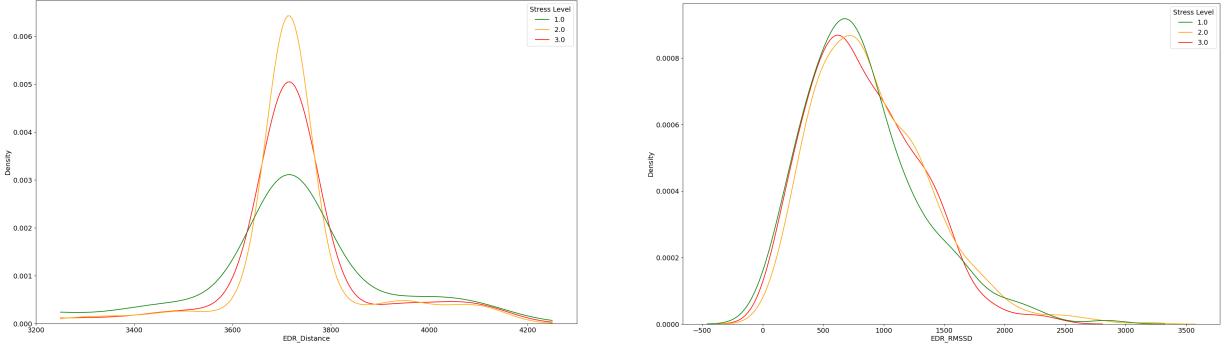


Figure 31: The EDR features kernel density plots.

Surprisingly, stress does not seem to impact EDR distance in the BrainPatch dataset. To investigate if breathing variability changes, the EDR RMSSD was plotted, however shows little variation also. The Spider-fear equivalent shows to vary slightly, see Figure A.11, however conversely to expectations, where breathing rate is higher during low stress.

The features that were visually most separable for each feature type were used to reclassify the tuned SVM and Random Forest models. The Naive Bayes model was not calculated since it did not perform well even when all the features were used. To evaluate the feature types independently, HRV, ECG and EDR features were classified in different combinations with the other two types.

Feature Set	SVM		RF		Average	
	Binary	Three-level	Binary	Three-level	Binary	Three-level
meanNN, minNN, maxNN, NN_RMSSD, pNN50, DFA	0.601	0.468	0.559	0.728	0.58	0.598
meanNN, minNN, maxNN, NN_RMSSD, pNN50, DFA, EDR	0.618	0.476	0.594	0.745	0.606	0.6105
meanNN, minNN, maxNN, NN_RMSSD, pNN50, DFA, P_Duration	0.653	0.466	0.677	0.742	0.665	0.604
meanNN, minNN, maxNN, NN_RMSSD, pNN50, DFA, FQ_10-30	0.726	0.527	0.808	0.824	0.767	0.6755
meanNN, minNN, maxNN, NN_RMSSD, pNN50, DFA, P_Duration, FQ_10-30, EDR	0.736	0.586	0.862	0.843	0.799	0.7145

Table 13: Traditional ML model accuracies with different features for HRV, ECG and EDR selected.

From the six most separable HRV features, the performance is moderate with a maximum of 72.8% accuracy achieved using the Random Forest model. The addition of the EDR feature improves accuracy for both by 2%, indicating that there some separability of the feature, shown in Figure 31. The P_Duration feature, does not significantly increase three-level classification, however drastically improves binary classification by an average of 8%. This is likely due to the high separability of low-stress in the feature, displayed in Figure 30.

4.7 Optimisation

Window Length

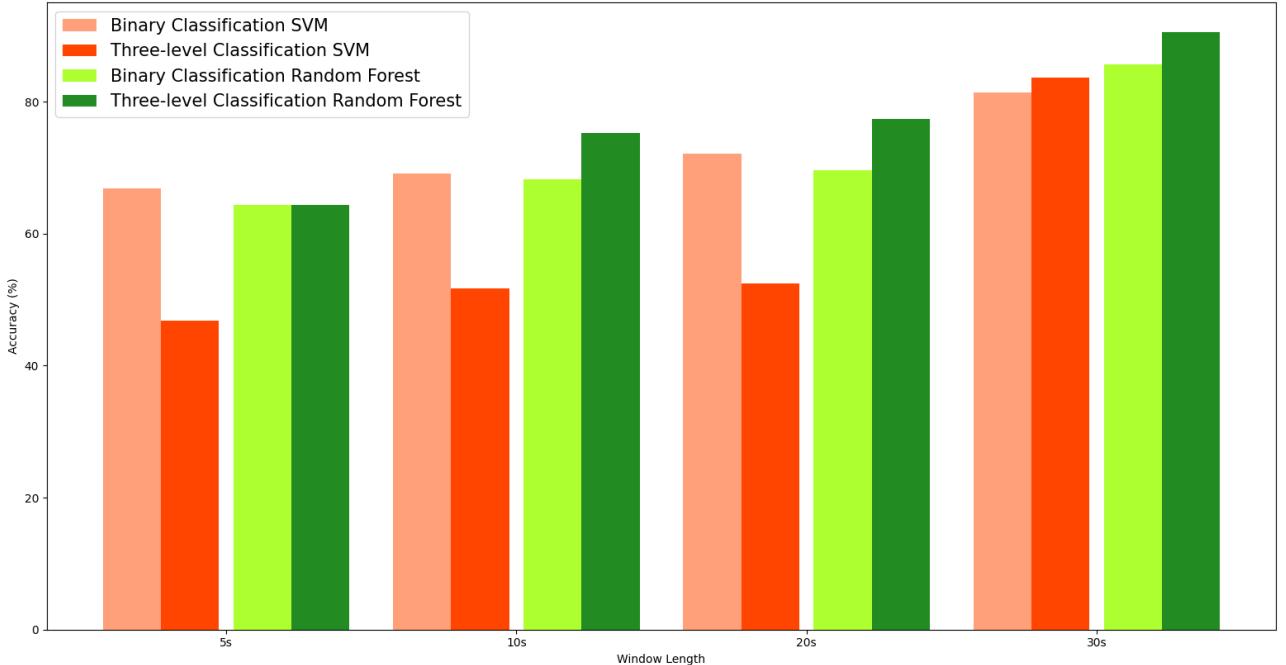


Figure 32: Comparison of different window lengths for Random Forests and SVM.

Shortening the window length shows to impact classification accuracy severely for both models, particularly from 30s to 20s. There is little loss in accuracy when lowering the window length from 20s to 10s for Random Forests achieving 68.2% and 75.2% for binary and three-level respectively.

Sampling Rate

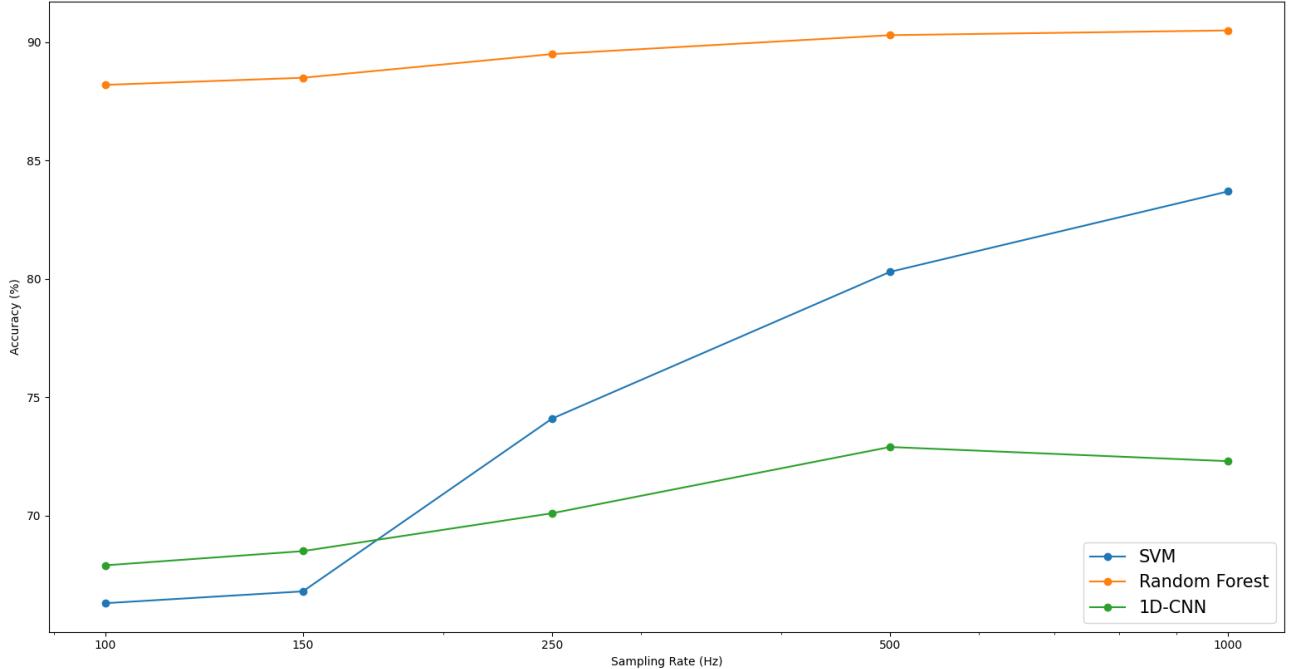


Figure 33: The effect of device sampling rate on accuracy for three-level classification.

Random Forests shows to experience a minor reduction in accuracy when the sampling rate is lowered, achieving 88.2% at 100Hz - only 2.3% lower than 1kHz. Conversely, SVM demonstrates to suffer, particularly below 500Hz. The CNN shows to improve by 0.6% from 1kHz to 500Hz, however this could be due to its non-deterministic nature, and after the accuracy declines linearly to 67.9% accuracy at 100Hz.

Sampling Rate (Hz)	Number of Parameters	Runtime Normalized to Random Forest Three-Level Classification
100	28,866	3.58
150	30,786	3.77
250	34,626	3.97
500	44,226	6.73
1000	63,426	11.21

Table 14: The effect of sampling rate on the 1D-CNN normalized runtime for three-level classification.

Since the number of neurons in the 1D CNN is dictated by the size of the feature vector and thus the number of samples, lowering the sampling rate drastically improves the model runtime. At 100Hz the sampling rate is now only 3.58 times larger than the Random Forest model.

5 Discussion

5.1 Device Performance

5.1.1 Signal Quality

From Figure 12, the raw ECG signal shows to attenuate noise extremely well in comparison to the Spider-fear device, see Figure A.8, which shows considerably more noise relative to its low frequency component. Despite the absence of a RLD electrode, the common mode rejection of the instrumentation amplifier and the high-pass filter effectively attenuate noise, as demonstrated. However, the ECG waveform itself undergoes significant distortion, as evident from the artifacts discussed in Artifact 3.3.1.

The isoelectric nature of the ECG (Artifact 3) resembles a V4 lead configuration of a chest ECG recording, whereby the ECG has comparably sized R and S waves, see Figure A.13. This results in a reduction in the signal-to-quantization-noise ratio since the rail-to-rail output fits to the large amplitude range and consequently results in a lowering of the relative amplitude of other waves, contributing to Artifact 1 and 2. Kennedy et Al. determined that to maximise the r-peak amplitude during bipolar chest ECG, electrodes should be placed at V2 and V5 [101]. This configuration is comparable to the Wellue device, discussed in Section 2.1. To continue using the chest strap for improved comfort, the dry electrodes could be placed further apart, which should lower the S peak amplitude, thereby increasing the relative amplitudes of the other waves, resulting in more information for the other peaks, that should improve both manual and automatic feature extraction.

5.1.2 Signal Reliability

The reliability of the signal was mixed, with most recordings showing consistent ECG traces that resulted in few segments falling outside of the healthy range during feature extraction. However, 4 out of 20 recordings had to be discarded due to severe artifacts in the signal, see Figure 13 for examples. This is likely due to poor electrode contact, which could be improved by tightening the strap or adding water to improve skin conduction, but likely at the compromise of user comfort. Another fix would be by utilising the leads-off detection system which detects if an electrode has lost its connection and could notify the user, to correct the electrode contact issue. Another reliability issue was where the signal conditioner output became partially or fully positively saturated. This is a common problem for biosensor front-ends, since the large gain of its differential amplifier can become saturated at the output if there is a non-negligible input offset voltage. The AD8232 implements a window comparator to detect saturation and restore the signal. However, since this does not omit saturation, it could be due to the leads off detection system, since the restoring circuitry is disabled if no electrodes appear to be present. More investigation into the circuitry will be required to fix this issue but does not appear to be a concrete limitation for chest strap ECG.

5.1.3 Feature Extraction

As shown in Section 3.4, the P, R, S and T waves can be detected accurately and reliably from the BrainPatch ECG signal, however the Q wave was unable to be detected. The lack of QT measurement therefore, resulted in the inability to measure VRV, a key stress signature that is explained in Section 1.4. As explained above, by adjusting the electrode placement, the Q wave should become more pronounced and could achieve similar results from that of the Spider-fear ECG. The EDR signal, could be obtained through the measurement of the R peak amplitudes, where the distance between maxima showed to be most reliable to extract respiration rate. The respiration rate was not compared with a respiration belt device and thus serves only as a demonstration of how additional stress-related features can be extracted from ECG.

5.2 Classification Performance

5.2.1 Model Comparison

Before evaluating classification performance, the abnormal improvement in accuracy from binary to three-level classification for SVM and Random Forest using the BrainPatch dataset must be addressed. This irregularity is likely due to overfitting in the binary classifier caused by insufficient training data. By adding additional data for three-level classification, through the medium samples and artificial samples from SMOTE, performance is significantly improved, thus indicating insufficient data for low and high stress. Insights from the classification findings can still be obtained but caution must be taken before making assertive claims.

Random Forests have demonstrated to be the most effective stress classification method for both datasets and classification types and adheres to the findings of Ihmig et Al., who also found it to perform over other traditional methods, including SVM. In particular, the model has shown to be the most effective classifier for medium stress, the most difficult to distinguish class. Moreover, for three-level classification, the runtime is comparable to SVM, making it a superior option in stress classification. However, SVM demonstrates exceptionally low runtime for binary classification and could offer a good compromise between performance and efficiency. Additionally, as shown in Krittawong et al., SVM could reach or surpass the accuracy of Random Forest with large ECG datasets.

The high performance of binary classification using Naive Bayes found in Keshan et Al. could not be replicated and is likely to be due to the large 5 minute window used in the study and only classifying on a per participant basis. However, it could be argued that the features used to achieve 100% binary classification accuracy in the study (average RR, SS and QQ durations) may require less computational power than that of models that perform using smaller window-samples but require feature extraction with more mathematical computation such as DFA. Thus Naive Bayes should not be dismissed as an option for lightweight stress classification, however may be limited due to its assumption that features are unrelated and follow a normal distribution.

The results of the 1D CNN has shown excellent classification accuracy for the spider-fear dataset. Three-level classification using five-fold cross validation has surpassed the highest accuracy recorded by Ihmig et al. by 4.5%, scoring an accuracy of 77.6%. This demonstrates the remarkable trait of the CNN - that an ultra-short window of 3 seconds can outperform the accuracy of a 60 second clip using both ECG and EDA features. Additionally, the run-time of the CNN showed to be only 3.58 times slower than that of Random Forest for a 100Hz signal and does not require feature extraction. Furthermore, Hwang et al. found that adding a recurrent neural network to the 1D CNN has shown to increase the accuracy of three-level classification, obtaining an accuracy of 3.84% more than that of Tzevelekakis et al. on the same database. This network is at the cost of 39,772 parameters compared to 28,866 and therefore further investigation must ascertain the feasibility of the networks for on-chip detection. The findings of the model using the BrainPatch dataset do not exceed the performance of the SVM and Random Forest models. This may be due to a lower temporal data from the BrainPatch's device caused by excessive filtering or from the reduction in signal resolution caused by the large S wave amplitude. More likely, it is due to the limited training data and thus must be re-evaluated once more data is collected.

5.2.2 Feature Set Comparison

When all the features detailed in Section 3.4 were used for classification, a three-level accuracy of 75.2% was achieved for the spider-fear dataset using a 30 second window. Comparatively, in their study a maximum of 73.4% was achieved using a 60 second window using the six most separable features both from ECG and EDA. Conversely, for binary classification the opposite occurred, whereby their study achieved 3.7% more than our findings, reaching an accuracy of 90.9%. The significant improvement in three-level classification demonstrates that the quantity of features is more pivotal to distinguishing subtle variations in stress. Nonetheless, EDA does provide more certainty to binary stress classification, and in the literature study Mohino et Al. [20], showed to benefit from smaller window-sizes (10s) rather

than 40s for ECG, thereby demonstrating its suitability for low-resource stress detection.

Determining the most effective feature sets for stress is imperative as it directly relates to the computational power required to classify traditional models, and thus dictates if on-chip detection is possible. From Figure 29-31, it was shown that the features extracted from HRV analysis yielded the most separability between classes. The mean, maximum, and minimum NN intervals were found to be highly effective in both classification tasks, aligning with the results reported by Ihmig et al. Additionally, RMSSD and pNN50 were observed to be superior to the standard deviation unlike their study. However, KDEs only provide qualitative evidence for the separability of features, thus the optimal features found in this study are not definitive.

Dimensionality reduction techniques, such as Principal Components, have not been investigated in this study however, have been shown to be a useful tool for ECG feature reduction [102]. Additionally, this study relies on the simplistic assumption that the entire feature set extracted will yield the best accuracy scores. Although a large reduction in features showed to reduce accuracy, some features in the set may have caused high overfitting and impacted accuracy in a negative way. More investigation is required to assure this is minimised and reiterates the need for careful selection of features for both accuracy and efficiency.

As stated above, the current device’s inability to measure VRV, is a key limitation to this study, and once this issue is addressed, further investigation should be pursued. Nonetheless, studies have shown QT detection to be unreliable and instead the use of RT intervals to map VRV [103, 104]. Additionally, as described in Section 1.4, Imam et Al. showed that EDR was able to distinguish ventricular repolarization dynamics, and improved binary stress classification of their dataset by approximately 12%. However, across both datasets examined in this study, the EDR signal only contributed to a 2% increase, aligning to the findings of Ihmig et Al. who neglected the RESP signal as an effective feature. It is unclear why there is such a large difference between the findings in these two studies for EDR, and more investigation must be conducted to validate its effectiveness. Regardless, for low-resource classification using the dataset of this study, there are other features that would improve accuracy more significantly, for example the P wave duration, which showed to contribute immensely to binary classification, increasing accuracy by 11.8% with Random Forest. Thus, ECG has definitively shown its superiority to only utilising HRV features extracted from a more simplistic heart rate sensor like PPG.

6 Limitations and Future Work

It is imperative to consider the limitations of the experimental design. Firstly, the number of participants are limited, and the imbalance between stress levels of these participants, shown in Table 4, could cause the models to associate a participant’s unique features with the most probabilistic stress level, leading to a higher accuracy. Secondly, the correction of the overall class imbalance in the training set using SMOTE, leads to an increase in non-deterministic results since it creates synthetic samples randomly [105], thereby increasing the deviation and lowering reliability. Retrospectively, Cohen’s Kappa statistic [106], utilised in Ihmig et Al., could have been implemented instead of SMOTE to prevent the need for imbalance correction, however this could have worsened model accuracy, particularly for the CNN due to its large training set requirement.

The significant challenge for the experimental design was how to induce and record stress level. Firstly, the participants reported facing difficulty in rating their stress levels, as distinguishing between medium and high stress was often challenging. Secondly, the experiment was designed to simulate work-induced stress that would be realistic to daily use. This approach stands in contrast to RCTs that typically employ protocols to induce stress, such as exposure to spider-clips or more commonly, stress tests like the Stroop Color Word test, bot of which should elicit a more consistent stress response. However, it has been demonstrated that conditions of active and passive mental stress, determined by the amount of control over the stressor, promote different levels of sympathetic and parasympathetic stimulation to the heart, thereby producing a different stress response [31]. Moreover, time-of-day was not accounted

for and HRV shows to have a lower frequency during the morning, since cortisol is at its highest [107]. In this study, active stress is assumed to be induced, however the stressor is uncontrolled and hence the stress response from the participants in the study may vary considerably and impact the results of this study.

Additional investigation must be conducted to identify whether the suitability of the different classification methods discussed in this study are suitable for on-chip detection. The key limitations for embedded stress detection systems are power consumption, computational performance and memory. If computational power is the limiting factor, the raw ECG can be sent via WiFi, which demonstrates to be reliable in home and work settings. However, for daily use this is unfeasible, and in addition, transmitting data with a high-sampling rate over WiFi resulted in a very limited battery life of 6.8hrs. Bluetooth low energy would significantly improve battery life, and should be investigated since the study shows that low frequency (100Hz) ECG still performs accurate stress detection. Another solution is on-chip feature extraction and classification to save on computational power. The use of analogue circuitry to detect features such as the QRS complex should not be overlooked, even in the age of digital signal processing. Circuitry that detect the QRS complex have existed since the 1980s [108] and still output new research with more energy-efficient techniques [109]. The device could then extract the 6 most effective HRV features from the R-R intervals and use a lightweight model to classify.

Keshan et Al. also highlighted the effectiveness of personalized individual stress analysis, whereby a low-stress sample ECG induced by a participant resting, could be used to distinguish that of high stress with 100% accuracy. Similarly, Gonzalez-Carabarin et al. [110] addressed inter-individual variability in stress responses by using individual classification methods and found that the personalised model was more effective than the general model, where lack of individual calibration was thought of as the limiting factor for achieving high reliability in stress classification. These methods are applied in a medical context however could be adapted to commercial devices.

7 Project Review and Conclusions

The project accomplished its objectives detailed in Section 1.2, and are summarised below:

1. **Research Aims:** The research conducted demonstrates the versatility of ECG for determining sympathetic activation levels and a comprehensive feature set that can be utilised during classification was obtained. The most proven ML models from other literature were selected, however this was not an exhaustive list, and several other models that were not implemented in this report could be investigated.
2. **Design Aims:** The development of the ECG device was conducted, with particular attention to the analogue front-end design. The cost of the device is minimal in comparison to that of a smartwatch, see Table A.6, however this is not inclusive of the stress detector itself, which may require large microcontroller processing power otherwise server infrastructure if on-chip detection is unfeasible. The signal quality from the device and digital signal processing, resulted in a clean, high fidelity signal, where all waveforms but the Q wave could be automatically detected. The IMU showed to function, however was only implemented in a primitive way to account for movement. Similarly, Wi-Fi streaming showed to stream data consistently and at a sufficiently high sampling rate for the required specification, however was not utilised in the study.
3. **Experimental Aims:** Detailed in Section 6, the data obtained from the study has several flaws therefore the results do not serve as concrete findings and must be researched in a more controlled manner. Nonetheless, the models were optimised and evaluated, leading to insights about the models researched, feature selection, and other factors such as window length and sample rate. More investigation will be required to evaluate if the runtimes of these models demonstrate feasibility for on-chip detection.

In conclusion, this study provides promising evidence for the use of bipolar chest strap ECG as a wearable and non-invasive method for stress classification. The elimination of disposable or wet electrodes greatly enhances wearability and convenience for users. Additionally, the superior classification power of ECG over PPG suggests the potential for more accurate and reliable stress detection. Random Forest and 1D CNN models demonstrate to be an effective approach to light-weight stress classification, which could potentially be integrated into an on-chip detection system in future research. Overall, this study contributes to the growing body of research on wearable technology for stress monitoring, and holds promise for practical applications for reducing the global stress epidemic.

References

- [1] Junxian Cai et al. "Multi-ECGNet for ECG Arrhythmia Multi-Label Classification". eng. In: *IEEE access* 8 (2020), pp. 110848–110858. ISSN: 2169-3536.
- [2] Amulya Agrawal et al. "ECG-iCOVIDNet: Interpretable AI model to identify changes in the ECG signals of post-COVID subjects". en. In: *Comput. Biol. Med.* 146.105540 (July 2022), p. 105540.
- [3] Hannah Ritchie, Fiona Spooner, and Max Roser. *Causes of death*. 2018. URL: <https://ourworldindata.org/causes-of-death>.
- [4] Geoffrey Verrall et al. "Influence of type and duration of training on the presence of an abnormal ECG in high-performance athletes". eng. In: *Heart Asia* 11.1 (2019), e011120–e011120. ISSN: 1759-1104.
- [5] Kyandoghere Kyamakya. *Emotion and Stress Recognition Related Sensors and Machine Learning Technologies*. eng. Basel, Switzerland: MDPI - Multidisciplinary Digital Publishing Institute, 2021.
- [6] Christina Maslach and Susan E. Jackson. "The measurement of experienced burnout". In: *Journal of Organizational Behavior* 2.2 (Apr. 1981), pp. 99–113. DOI: 10.1002/job.4030020205. URL: <https://doi.org/10.1002/job.4030020205>.
- [7] Christina Maslach, Susan E. Jackson, and Michael P. Leiter. "Maslach Burnout Inventory: Third edition." In: Evaluating stress: A book of resources. Lanham, MD, US: Scarecrow Education, 1997, pp. 191–218. ISBN: 0-8108-3231-3 (Hardcover).
- [8] *Burnout*. <https://mentalhealth-uk.org/burnout/>. Accessed: 2023-04-02. 2020.
- [9] Mohd. Razali Salleh. "Life event, stress and illness". In: *The Malaysian journal of medical sciences* 15.4 (2009). ISSN: 1394-195X.
- [10] "COVAX delivers its 1 billionth COVID-19 vaccine dose. COVID-19 pandemic triggers 25 Percent increase in prevalence of anxiety and depression worldwide". eng. In: *Saudi medical journal* 43.4 (2022), pp. 438–439. ISSN: 0379-5284.
- [11] Brent D. Winslow et al. "Automated stress detection using mobile application and wearable sensors improves symptoms of mental health disorders in military personnel". In: *Frontiers in Digital Health* 4 (Aug. 2022). DOI: 10.3389/fdgth.2022.919626. URL: <https://doi.org/10.3389/fdgth.2022.919626>.
- [12] Florane Pasquier et al. "Impact of Galvanic Vestibular Stimulation on Anxiety Level in Young Adults". In: *Frontiers in Systems Neuroscience* 13 (Apr. 2019). DOI: 10.3389/fnsys.2019.00014. URL: <https://doi.org/10.3389/fnsys.2019.00014>.
- [13] L. Rochette and C. Vergely. "Hans Selye and the stress response: 80 years after his "letter" to the Editor of Nature". In: *Annales de Cardiologie et d'Angéiologie* 66.4 (Sept. 2017), pp. 181–183. DOI: 10.1016/j.ancard.2017.04.017. URL: <https://doi.org/10.1016/j.ancard.2017.04.017>.
- [14] HANS SELYE. "A Syndrome produced by Diverse Nocuous Agents". In: *Nature* 138.3479 (July 1936), pp. 32–32. DOI: 10.1038/138032a0. URL: <https://doi.org/10.1038/138032a0>.
- [15] R. Murison. *Neuroscience of Pain, Stress, and Emotion*. Elsevier, 2016. DOI: 10.1016/c2013-0-16065-5. URL: <https://doi.org/10.1016/c2013-0-16065-5>.
- [16] A Sanchez. "Changes in norepinephrine and epinephrine concentrations in adrenal gland of the rats submitted to acute immobilization stress". In: *Pharmacological Research* 48.6 (Dec. 2003), pp. 607–613. DOI: 10.1016/s1043-6618(03)00241-x. URL: [https://doi.org/10.1016/s1043-6618\(03\)00241-x](https://doi.org/10.1016/s1043-6618(03)00241-x).
- [17] Michael D. Breed and Janice Moore. "Neurobiology and Endocrinology for Animal Behaviorists". In: *Animal Behavior*. Elsevier, 2012, pp. 25–65. DOI: 10.1016/b978-0-12-372581-3.00002-7. URL: <https://doi.org/10.1016/b978-0-12-372581-3.00002-7>.
- [18] *Parasympathetic Nervous System - An Overview*. <https://www.sciencedirect.com/topics/computer-science/parasympathetic-nervous-system>. Accessed: 2023-04-05. 2023.
- [19] Cheyenne Samson and Ahyeon Koh. "Stress Monitoring and Recent Advancements in Wearable Biosensors". In: *Frontiers in Bioengineering and Biotechnology* 8 (Sept. 2020). DOI: 10.3389/fbioe.2020.01037. URL: <https://doi.org/10.3389/fbioe.2020.01037>.

- [20] Inma Mohino-Herranz et al. “Activity Recognition Using Wearable Physiological Measurements: Selection of Features from a Comprehensive Literature Study”. In: *Sensors* 19.24 (Dec. 2019), p. 5524. DOI: 10.3390/s19245524. URL: <https://doi.org/10.3390/s19245524>.
- [21] Eduardo Perez-Valero et al. “Quantitative Assessment of Stress Through EEG During a Virtual Reality Stress-Relax Session”. In: *Frontiers in Computational Neuroscience* 15 (July 2021). DOI: 10.3389/fncom.2021.684423. URL: <https://doi.org/10.3389/fncom.2021.684423>.
- [22] William Powell. “Design and Implementation of a Chest Strap ECG Recorder with Integrated IMU for Decoding Stress using Machine Learning Methods”. In: (2023).
- [23] Yun Liu and Siqing Du. “Psychological stress level detection based on electrodermal activity”. In: *Behavioural Brain Research* 341 (Apr. 2018), pp. 50–53. DOI: 10.1016/j.bbr.2017.12.021. URL: <https://doi.org/10.1016/j.bbr.2017.12.021>.
- [24] Adam Gacek and Witold Pedrycz. *ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence*. ger ; eng. 1. Aufl. Vol. 9780857298683. London: Springer Verlag London Limited, 2011. ISBN: 9780857298676.
- [25] Teresa Henriques et al. “Nonlinear Methods Most Applied to Heart-Rate Time Series: A Review”. In: *Entropy* 22.3 (Mar. 2020), p. 309. DOI: 10.3390/e22030309. URL: <https://doi.org/10.3390/e22030309>.
- [26] R. Castaldo et al. “Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis”. In: *Biomedical Signal Processing and Control* 18 (Apr. 2015), pp. 370–377. DOI: 10.1016/j.bspc.2015.02.012. URL: <https://doi.org/10.1016/j.bspc.2015.02.012>.
- [27] Jakub Parak et al. “Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor”. eng. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Vol. 2015-. United States: IEEE, 2015, pp. 8099–8102. ISBN: 9781424492718.
- [28] R. Almeida et al. “QT variability and HRV interactions in ECG: quantification and reliability”. In: *IEEE Transactions on Biomedical Engineering* 53.7 (2006), pp. 1317–1329. DOI: 10.1109/TBME.2006.873682.
- [29] M. H. Imam et al. “Effect of ECG-derived respiration (EDR) on modeling ventricular repolarization dynamics in different physiological and psychological conditions”. In: *Medical & Biological Engineering & Computing* 52.10 (Aug. 2014), pp. 851–860. DOI: 10.1007/s11517-014-1188-0. URL: <https://doi.org/10.1007/s11517-014-1188-0>.
- [30] *QT Interval - An Overview*. <https://www.sciencedirect.com/topics/neuroscience/qt-interval>. Accessed: 2023-04-05. 2023.
- [31] Gábor Andrászky et al. “Mental Stress May Induce QT-Interval Prolongation and T-Wave Notching”. In: *Annals of Noninvasive Electrocardiology* 12.3 (July 2007), pp. 251–259. DOI: 10.1111/j.1542-474x.2007.00169.x. URL: <https://doi.org/10.1111/j.1542-474x.2007.00169.x>.
- [32] Lauri Toivonen, Kirsi Helenius, and Matti Viitasalo. “Electrocardiographic Repolarization During Stress From Awakening on Alarm Call”. In: *Journal of the American College of Cardiology* 30.3 (July 1997), pp. 774–779. DOI: 10.1016/s0735-1097(97)00222-2. URL: [https://doi.org/10.1016/s0735-1097\(97\)00222-2](https://doi.org/10.1016/s0735-1097(97)00222-2).
- [33] Petri Haapalahti et al. “Effects of cardiovascular autonomic function tests on QT dispersion in the 12-lead electrocardiogram of healthy patients”. In: *Journal of Electrocardiology* 33.4 (Oct. 2000), pp. 321–327. DOI: 10.1054/jelc.2000.18357. URL: <https://doi.org/10.1054/jelc.2000.18357>.
- [34] Randall Steven Jorgensen. “Respiratory Sinus Arrhythmia”. eng. In: *Encyclopedia of Behavioral Medicine*. Cham: Springer International Publishing, 2020, pp. 1910–1910. ISBN: 9783030399016.
- [35] Paul Grossman and Edwin W. Taylor. “Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions”. In: *Biological Psychology* 74.2 (2007), pp. 263–285. DOI: 10.1016/j.biopsych.2005.11.014. URL: <https://doi.org/10.1016/j.biopsych.2005.11.014>.

- [36] L. Zhao, S. Reisman, and T. Findley. “Derivation of respiration from electrocardiogram during heart rate variability studies”. In: *Computers in Cardiology 1994*. 1994, pp. 53–56. DOI: 10.1109/CIC.1994.470251.
- [37] Surita Sarkar, Saurabh Pal, and Parthasarathi Bhattacharyya. “A Comparative Analysis Between EDR and Respiration Signal: A Pilot Study with Normal Subjects”. In: *Advances in Intelligent Systems and Computing*. Springer International Publishing, Oct. 2018, pp. 595–604. DOI: 10.1007/978-3-319-74808-5_52. URL: https://doi.org/10.1007/978-3-319-74808-5_52.
- [38] Jennifer A Healey and Rosalind W Picard. *Stress Recognition in Automobile Drivers*. 2008. DOI: 10.13026/C2SG6B. URL: <https://physionet.org/content/drivedb/>.
- [39] M. Abbod. “Advanced Signal Processing in Wearable Sensors for Health Monitoring”. In: (2022). URL: https://mdpi-res.com/bookfiles/book/5368/Advanced_Signal_Processing_in_Wearable_Sensors_for_Health_Monitoring.pdf?v1678372043.
- [40] Chayakrit Krittawong et al. “Artificial Intelligence in Precision Cardiovascular Medicine”. In: *Journal of the American College of Cardiology* 69.21 (2017), pp. 2657–2664. ISSN: 0735-1097. DOI: <https://doi.org/10.1016/j.jacc.2017.03.571>. URL: <https://www.sciencedirect.com/science/article/pii/S0735109717368456>.
- [41] Igor Stancin et al. “EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization”. In: *Sensors* 21.20 (2021). ISSN: 1424-8220. DOI: 10.3390/s21206932. URL: <https://www.mdpi.com/1424-8220/21/20/6932>.
- [42] A. Khazaee and A. Ebrahimzadeh. “Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features”. In: *Biomedical Signal Processing and Control* 5.4 (2010), pp. 252–263. ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2010.07.006>. URL: <https://www.sciencedirect.com/science/article/pii/S1746809410000546>.
- [43] Richa Gupta, M. Afshar Alam, and Parul Agarwal. “Modified Support Vector Machine for Detecting Stress Level Using EEG Signals”. In: *Computational Intelligence and Neuroscience* 2020 (Aug. 2020), pp. 1–14. DOI: 10.1155/2020/8860841. URL: <https://doi.org/10.1155/2020/8860841>.
- [44] Prince Kumar, Shruti Garg, and Ashwani Garg. “Assessment of Anxiety, Depression and Stress using Machine Learning Models”. In: *Procedia Computer Science* 171 (2020). Third International Conference on Computing and Network Communications (CoCoNet’19), pp. 1989–1998. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2020.04.213>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050920311984>.
- [45] Feng-Tso Sun et al. “Activity-Aware Mental Stress Detection Using Physiological Sensors”. In: *Mobile Computing, Applications, and Services*. Ed. by Martin Gris and Guang Yang. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 211–230. ISBN: 978-3-642-29336-8.
- [46] Yu-Hung Chuang et al. “Automatic Classification of Myocardial Infarction Using Spline Representation of Single-Lead Derived Vectorcardiography”. In: *Sensors* 20.24 (2020). ISSN: 1424-8220. DOI: 10.3390/s20247246. URL: <https://www.mdpi.com/1424-8220/20/24/7246>.
- [47] Konstantinos Tzevelekakis, Zinovia Stefanidi, and George Margetis. “Real-Time Stress Level Feedback from Raw Ecg Signals for Personalised, Context-Aware Applications Using Lightweight Convolutional Neural Network Architectures”. In: *Sensors* 21.23 (2021). ISSN: 1424-8220. DOI: 10.3390/s21237802. URL: <https://www.mdpi.com/1424-8220/21/23/7802>.
- [48] N. Keshan, P. V. Parimi, and I. Bichindaritz. “Machine learning for stress detection from ECG signals in automobile drivers”. In: *2015 IEEE International Conference on Big Data (Big Data)*. 2015, pp. 2661–2669. DOI: 10.1109/BigData.2015.7364066.
- [49] “Advanced Analytical Theory and Methods”. eng. In: *Data Science & Big Data Analytics*. Indianapolis, Indiana: John Wiley & Sons, Inc, 2015. ISBN: 111887613X.
- [50] *Feature Crosses: Encoding Nonlinearity*. <https://developers.google.com/machine-learning/crash-course/feature-crosses/encoding-nonlinearity>. Accessed: 2023-04-22. 2022.
- [51] Pratap Dangeti. *Statistics for Machine Learning*. eng. Birmingham: Packt Publishing, 2017. ISBN: 9781788291224.

- [52] Grace Zhang. *What is the kernel trick? Why is it important?* <https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d>. Accessed: 2023-04-20. 2018.
- [53] Frank R. Ihmig et al. “On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals”. In: *PLOS ONE* 15.6 (June 2020). Ed. by Paweł Pławiak, e0231517. DOI: 10.1371/journal.pone.0231517. URL: <https://doi.org/10.1371/journal.pone.0231517>.
- [54] *A Beginner’s Guide to Neural Networks and Deep Learning*. 2023. URL: <http://wiki.pathmind.com/neural-network>.
- [55] Özal Yıldırım et al. “Arrhythmia detection using deep convolutional neural network with long duration ECG signals”. In: *Computers in Biology and Medicine* 102 (2018), pp. 411–420. ISSN: 0010-4825. DOI: <https://doi.org/10.1016/j.combiomed.2018.09.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0010482518302713>.
- [56] Yunan Wu et al. *A Comparison of 1-D and 2-D Deep Convolutional Neural Networks in ECG Classification*. 2018. arXiv: 1810.07088 [cs.CV].
- [57] Dan Li et al. “Classification of ECG signals based on 1D convolution neural network”. In: *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. 2017, pp. 1–6. DOI: 10.1109/HealthCom.2017.8210784.
- [58] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. “Personalized Monitoring and Advance Warning System for Cardiac Arrhythmias”. In: *Scientific Reports* 7.1 (Aug. 2017). DOI: 10.1038/s41598-017-09544-z. URL: <https://doi.org/10.1038/s41598-017-09544-z>.
- [59] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. “Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks”. In: *IEEE Transactions on Biomedical Engineering* 63.3 (2016), pp. 664–675. DOI: 10.1109/TBME.2015.2468589.
- [60] Bosun Hwang et al. “Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals”. In: *TELEMEDICINE and e-HEALTH* 24.10 (2018), pp. 753–772.
- [61] Sharif Abuadbba et al. “Can We Use Split Learning on 1D CNN Models for Privacy Preserving Training?” In: Oct. 2020, pp. 305–318. DOI: 10.1145/3320269.3384740.
- [62] Pérez-Enciso and Laura Zingaretti. “A Guide for Using Deep Learning for Complex Trait Genomic Prediction”. In: *Genes* 10 (July 2019), p. 553. DOI: 10.3390/genes10070553.
- [63] Vincent Dumoulin and Francesco Visin. *A guide to convolution arithmetic for deep learning*. 2018. arXiv: 1603.07285 [stat.ML].
- [64] Era Ajdaraga and Marjan Gusev. “Analysis of sampling frequency and resolution in ECG signals”. In: *2017 25th Telecommunication Forum (TELFOR)*. 2017, pp. 1–4. DOI: 10.1109/TELFOR.2017.8249438.
- [65] Kar Fye Alvin Lee et al. “Lowering the Sampling Rate: Heart Rate Response during Cognitive Fatigue”. In: *Biosensors* 12.5 (May 2022), p. 315. DOI: 10.3390/bios12050315. URL: <https://doi.org/10.3390/bios12050315>.
- [66] Ohhwan Kwon et al. “Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis”. In: *Healthcare Informatics Research* 24.3 (2018), p. 198. DOI: 10.4258/hir.2018.24.3.198. URL: <https://doi.org/10.4258/hir.2018.24.3.198>.
- [67] G.P. Pizzuti, S. Cifaldi, and G. Nolfe. “Digital sampling rate and ECG analysis”. In: *Journal of Biomedical Engineering* 7.3 (1985), pp. 247–250. ISSN: 0141-5425. DOI: [https://doi.org/10.1016/0141-5425\(85\)90027-5](https://doi.org/10.1016/0141-5425(85)90027-5). URL: <https://www.sciencedirect.com/science/article/pii/0141542585900275>.
- [68] Jonathan James Hyett Bray et al. “Single-lead ECGs (AliveCor) are a feasible, cost-effective and safer alternative to 12-lead ECGs in community diagnosis and monitoring of atrial fibrillation”. In: *BMJ Open Quality* 10.1 (Mar. 2021), e001270. DOI: 10.1136/bmjoq-2020-001270. URL: <https://doi.org/10.1136/bmjoq-2020-001270>.
- [69] Marcelle Schaffarczyk et al. “Validity of the Polar H10 Sensor for Heart Rate Variability Analysis during Resting State and Incremental Exercise in Recreational Men and Women”. In:

- Sensors* 22.17 (2022). ISSN: 1424-8220. DOI: 10.3390/s22176536. URL: <https://www.mdpi.com/1424-8220/22/17/6536>.
- [70] Katrina Hinde, Graham White, and Nicola Armstrong. “Wearable Devices Suitable for Monitoring Twenty Four Hour Heart Rate Variability in Military Populations”. In: *Sensors* 21.4 (2021). ISSN: 1424-8220. DOI: 10.3390/s21041061. URL: <https://www.mdpi.com/1424-8220/21/4/1061>.
- [71] Rahel Gilgen-Ammann, Theresa Schweizer, and Thomas Wyss. “RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise”. In: *European Journal of Applied Physiology* 119.7 (Apr. 2019), pp. 1525–1532. DOI: 10.1007/s00421-019-04142-5. URL: <https://doi.org/10.1007/s00421-019-04142-5>.
- [72] Andre Matthias Müller et al. “Heart Rate Measures From Wrist-Worn Activity Trackers in a Laboratory and Free-Living Setting: Validation Study”. In: *JMIR mHealth and uHealth* 7.10 (Oct. 2019), e14120. DOI: 10.2196/14120. URL: <https://doi.org/10.2196/14120>.
- [73] Tomáš Skála et al. “Feasibility of evaluation of Polar H10 chest-belt ECG in patients with a broad range of heart conditions”. In: *Cor et Vasa* 64.4 (Sept. 2022), pp. 411–422. DOI: 10.33678/cor.2022.083. URL: <https://doi.org/10.33678/cor.2022.083>.
- [74] Peter Kinget Chia-Hung Chen Shi-Gun Pan. “ECG Measurement System”. In: (2014).
- [75] *High-impedance and low-noise op amps enable dry electrodes in medical systems*. URL: <https://www.analog.com/en/design-notes/highimpedance-and-lownoise-op-amps-enable-dry-electrodes-in-medical-systems.html>.
- [76] Leila Safari et al. “Principles of Instrumentation Amplifiers”. eng. In: *Current-Mode Instrumentation Amplifiers*. Analog Circuits and Signal Processing. Cham: Springer International Publishing, 2019, pp. 1–13. ISBN: 9783030013424.
- [77] Vijay Singh Rathore et al. “Analog Front-End Design and Construction for ECG Monitoring System”. eng. In: *Rising Threats in Expert Applications and Solutions*. Vol. 1187. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore Pte. Limited, 2020, pp. 695–703. ISBN: 9811560137.
- [78] Gustavo Lenis et al. “Comparison of Baseline Wander Removal Techniques considering the Preservation of ST Changes in the Ischemic ECG: A Simulation Study”. eng. In: *Computational and mathematical methods in medicine* 2017 (2017), pp. 9295029–13. ISSN: 1748-670X.
- [79] Sonu Goel, G. Kau, and P. Toma. “A novel technique for stress recognition using ECG signal pattern”. In: *Current Pediatric Research* 21 (Jan. 2017), pp. 674–679.
- [80] Venkatesh Acharya. *Improving Common-Mode Rejection Using the Right-Leg Drive Amplifier*. 2017.
- [81] A. C. MettingVanRijn, A. Peper, and C. A. Grimbergen. “Amplifiers for bioelectric events: A design with a minimal number of parts”. In: *Medical & Biological Engineering & Computing* 32.3 (May 1994), pp. 305–310. DOI: 10.1007/bf02512527. URL: <https://doi.org/10.1007/bf02512527>.
- [82] Gaetano D. Gargiulo et al. “On the einthoven triangle: A critical analysis of the single rotating dipole hypothesis”. eng. In: 18.7 (2018), p. 2353. ISSN: 1424-8220.
- [83] Min Suk Lee et al. “Characterization of Ag/AgCl Dry Electrodes for Wearable Electrophysiological Sensing”. In: *Frontiers in Electronics* 2 (Jan. 2022). DOI: 10.3389/felec.2021.700363. URL: <https://doi.org/10.3389/felec.2021.700363>.
- [84] “Polar H10 Heart Rate Sensor System”. In: (Jan. 2019). URL: <https://www.polar.com/sites/default/files/static/science/white-papers/polar-h10-heart-rate-sensor-white-paper.pdf>.
- [85] Slavomir Kardos, Peter Balog, and Stanislav Slosarcik. “Gait dynamics sensing using IMU sensor array system”. eng. In: *Advances in electrical and electronic engineering* 15.1 (2017), pp. 71–76. ISSN: 1336-1376.
- [86] Sean Walton. *Linux socket programming*. eng. 1st edition. Place of publication not identified: Sams, 2001.

- [87] Frank R Ihmig et al. *Electrocardiogram, skin conductance and respiration from spider-fearful individuals watching spider video clips*. 2020. DOI: 10.13026/SQ6Q-ZG04. URL: <https://physionet.org/content/ecg-spider-clip/1.0.0/>.
- [88] Dominique Makowski et al. “NeuroKit2: A Python toolbox for neurophysiological signal processing”. In: *Behavior Research Methods* 53.4 (2021), pp. 1689–1696. DOI: 10.3758/s13428-020-01516-y. URL: <https://doi.org/10.3758%5C%2Fs13428-020-01516-y>.
- [89] Jiapu Pan and Willis J. Tompkins. “A Real-Time QRS Detection Algorithm”. eng. In: *IEEE transactions on biomedical engineering* BME-32.3 (1985), pp. 230–236. ISSN: 0018-9294.
- [90] Paul Kligfield et al. “Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: The Electrocardiogram and Its Technology A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology”. In: *Journal of the American College of Cardiology* 49.10 (2007), pp. 1109–1127. ISSN: 0735-1097. DOI: <https://doi.org/10.1016/j.jacc.2007.01.024>. URL: <https://www.sciencedirect.com/science/article/pii/S073510970700232X>.
- [91] Lei Cao et al. “A Novel Deep Learning Method Based on an Overlapping Time Window Strategy for Brain–Computer Interface-Based Stroke Rehabilitation”. In: *Brain Sciences* 12.11 (Nov. 2022), p. 1502. DOI: 10.3390/brainsci12111502. URL: <https://doi.org/10.3390/brainsci12111502>.
- [92] *Machine Learning Crash Course with TensorFlow APIs*. 2023. URL: <https://developers.google.com/machine-learning/crash-course>.
- [93] Tam Pham et al. “Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial”. In: *Sensors* 21.12 (2021). ISSN: 1424-8220. DOI: 10.3390/s21123998. URL: <https://www.mdpi.com/1424-8220/21/12/3998>.
- [94] Ankit Maheshwari et al. “Relation of Prolonged P-Wave Duration to Risk of Sudden Cardiac Death in the General Population (from the Atherosclerosis Risk in Communities Study)”. In: *The American Journal of Cardiology* 119.9 (May 2017), pp. 1302–1306. DOI: 10.1016/j.amjcard.2017.01.012. URL: <https://doi.org/10.1016/j.amjcard.2017.01.012>.
- [95] Bernd Porr. “A new R-peak detector stress test combining temporal jitter and accuracy (JA) reveals significant performance differences amongst popular detectors”. In: (Aug. 2019). DOI: 10.1101/722397. URL: <https://doi.org/10.1101/722397>.
- [96] ECG & ECHO Learning. *ECG interpretation: Characteristics of the normal ECG (P-wave, QRS complex, ST segment, T-wave)*. <https://ecgwaves.com/topic/ecg-normal-p-wave-qrs-complex-st-segment-t-wave-j-point/>. Accessed May 5, 2023. 2023.
- [97] Yasir Sattar and Lovish Chhabra. *Electrocardiogram*. StatPearls [Internet]. Jan. 2023. URL: <https://www.ncbi.nlm.nih.gov/books/NBK549803/>.
- [98] Ciprian Rezus et al. “QT interval variations and mortality risk: Is there any relationship?” In: *Anadolu Kardiyoloji Dergisi/The Anatolian Journal of Cardiology* 15.3 (Mar. 2015), pp. 255–258. DOI: 10.5152/akd.2015.5875. URL: <https://doi.org/10.5152/akd.2015.5875>.
- [99] Paul van Gent et al. “HeartPy: A novel heart rate algorithm for the analysis of noisy signals”. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 66 (2019), pp. 368–378. ISSN: 1369-8478. DOI: <https://doi.org/10.1016/j.trf.2019.09.015>. URL: <https://www.sciencedirect.com/science/article/pii/S1369847818306740>.
- [100] Pramod Gupta. “Introduction to machine learning in the cloud with Python : concepts and practices”. eng. In: Cham, Switzerland: Springer, 2021, p. 44. ISBN: 3-030-71270-2.
- [101] Alan Kennedy et al. “Optimisation of electrode placement for new ambulatory ECG monitoring devices”. eng. In: *Computing in Cardiology*. Vol. 43. 2016, pp. 101–104. ISBN: 9781509008964.
- [102] Lan-lan Chen et al. “Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers”. In: *Expert Systems with Applications* 85 (2017), pp. 279–291. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2017.01.040>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417417300477>.

- [103] Alberto Porta et al. “Complexity and Nonlinearity in Short-Term Heart Period Variability: Comparison of Methods Based on Local Nonlinear Prediction”. In: *IEEE Transactions on Biomedical Engineering* 54.1 (2007), pp. 94–106. DOI: 10.1109/TBME.2006.883789.
- [104] Alberto Porta et al. “RT variability unrelated to heart period and respiration progressively increases during graded head-up tilt”. In: *American Journal of Physiology-Heart and Circulatory Physiology* (2010).
- [105] Fredy Rodriguez-Torres. “Deterministic oversampling methods based on SMOTE.” eng. In: *Journal of Intelligent & Fuzzy Systems* 36.5 (2019), pp. 4945–4956. ISSN: 10641246.
- [106] Arie Ben-David. “Comparison of classification accuracy using Cohen’s Weighted Kappa”. In: *Expert Systems with Applications* 34.2 (2008), pp. 825–832. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2006.10.022>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417406003435>.
- [107] Y. Yamasaki et al. “Diurnal heart rate variability in healthy subjects: effects of aging and sex difference”. In: *American Journal of Physiology-Heart and Circulatory Physiology* 271.1 (July 1996), H303–H310. DOI: 10.1152/ajpheart.1996.271.1.h303. URL: <https://doi.org/10.1152/ajpheart.1996.271.1.h303>.
- [108] J. Fraden and M. R. Neuman. “QRS wave detection”. In: *Medical & Biological Engineering & Computing* 18.2 (Mar. 1980), pp. 125–132. DOI: 10.1007/bf02443287. URL: <https://doi.org/10.1007/bf02443287>.
- [109] Farnaz Morshedlou, Ali Asghar Orouji, and Nassim Ravanshad. “An energy-efficient analog circuit for detecting QRS complexes from ECG signal”. In: *Integration* 88 (2023), pp. 390–399. ISSN: 0167-9260. DOI: <https://doi.org/10.1016/j.vlsi.2022.11.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0167926022001481>.
- [110] L. Gonzalez-Carabarin et al. “Machine Learning for personalised stress detection: Inter-individual variability of EEG-ECG markers for acute-stress response”. In: *Computer Methods and Programs in Biomedicine* 209 (Sept. 2021), p. 106314. DOI: 10.1016/j.cmpb.2021.106314. URL: <https://doi.org/10.1016/j.cmpb.2021.106314>.
- [111] *ECG (EKG) Interpretation*. <https://oxfordmedicaleducation.com/ecgs/ecg-interpretation/>. Accessed: 2023-04-02. 2020.

A Appendix

Board Schematics

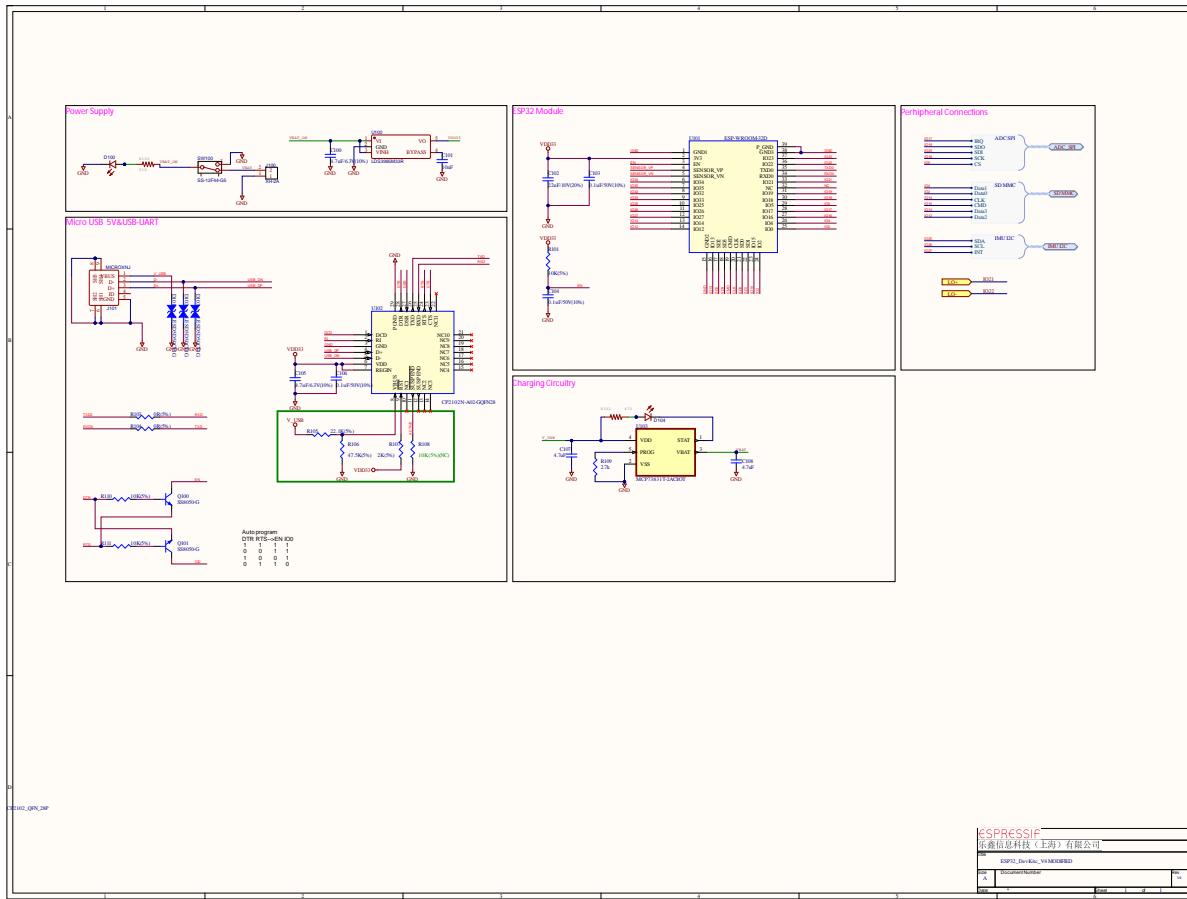


Figure A.1: ESP32 and required peripherals: USB-UART, power supply and charging circuitry.

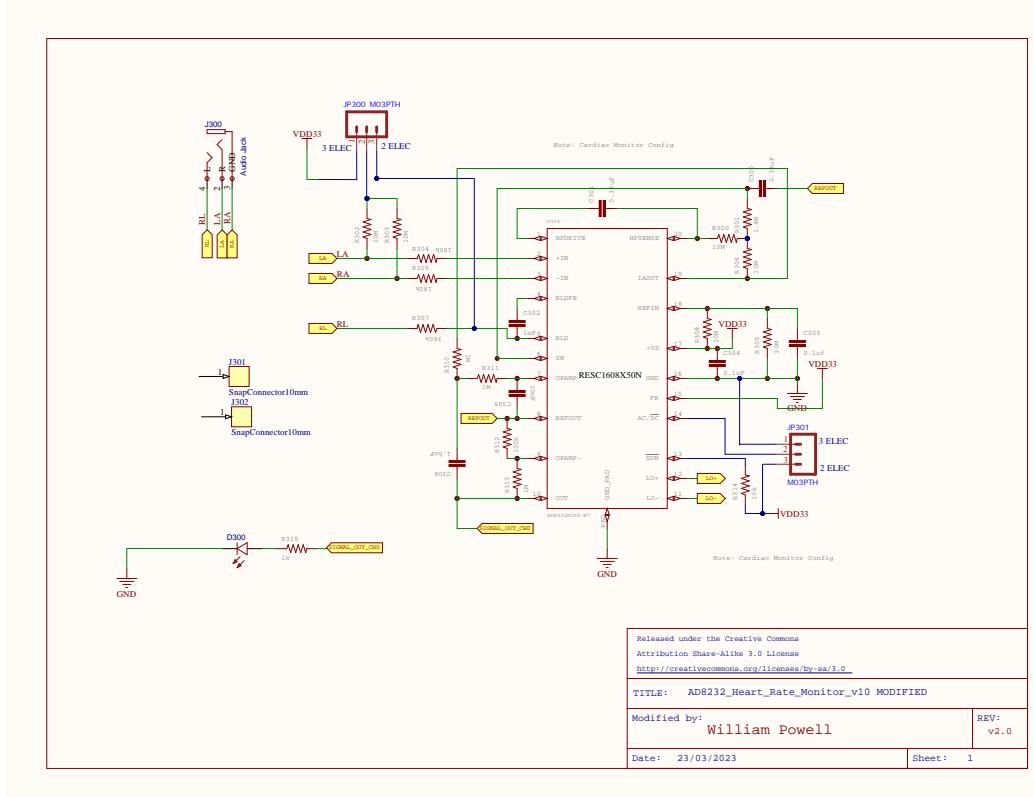


Figure A.2: AD8232 Signal Conditioner Circuitry

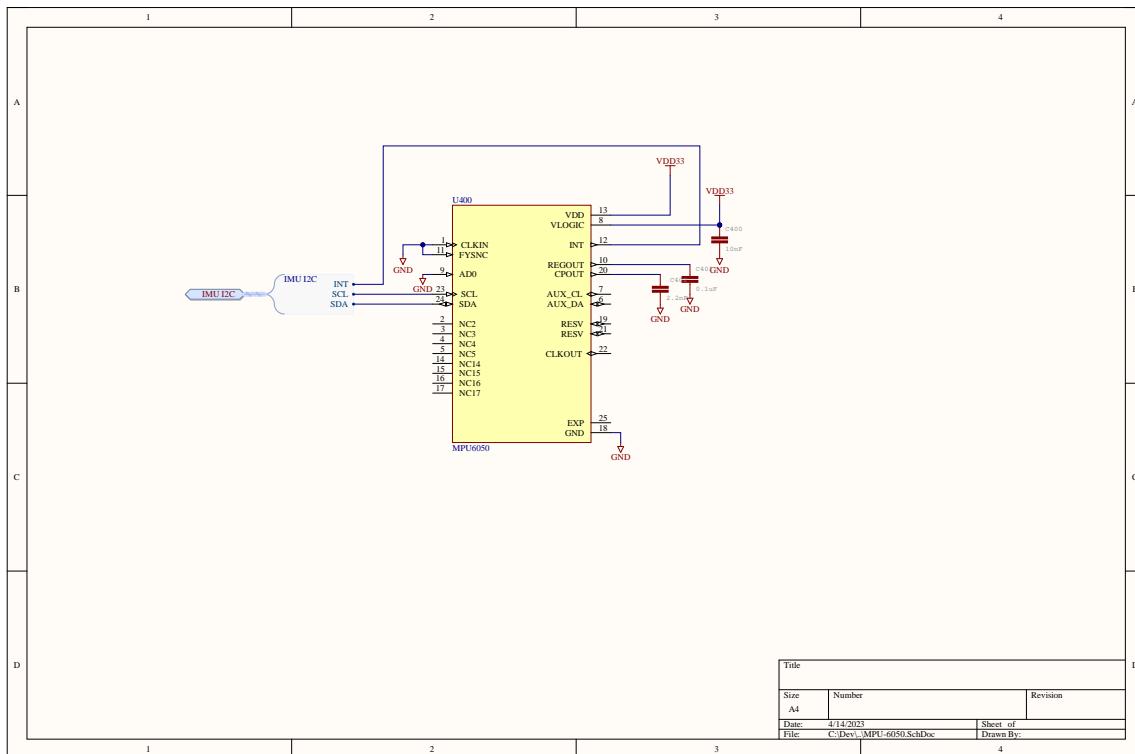


Figure A.3: MPU6050 Accelerometer and Gyroscope Circuitry

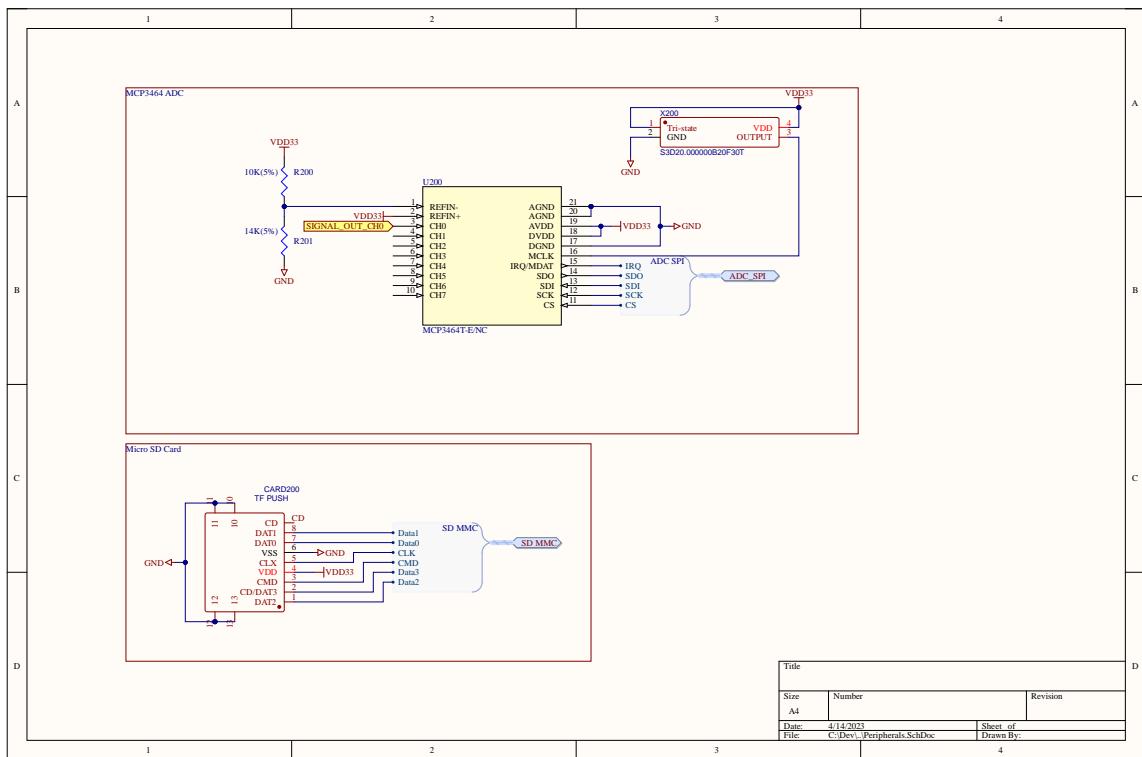


Figure A.4: MCP3464 ADC with external 20MHz clock and Micro SD Card.

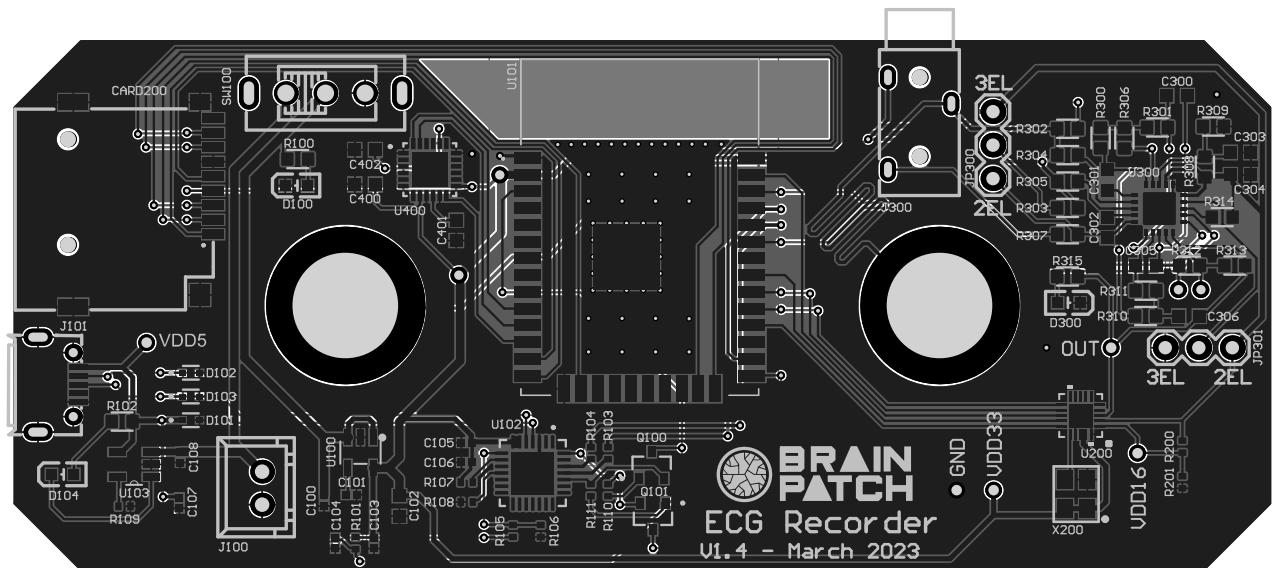


Figure A.5: PCB Board

Bill of Materials

Comment	Designator	JLCPCB Part #
0.33uF	C300,C301	C1615
10K(5%)(NC)	R101,R108, R110,R111,R200,R201	C269674
10nF	C305,C400	C57112
XH-2A	J100	C20079
2.2nF	C402	C1604
1k	R314	C2889391
20MHzOscillator	X200	C337680
ESP-WROOM-32D	U101	C473012
AudioJack	J300	C2689690
0R(5%)	R103,R104	C325377
180k	R304,R305	C267263
MICROXNJ	J101	C404969
SS8050-G	Q100,Q101	C164885
1nF	C302	C1588
0.1uf	C303,C304,C401	C14663
100k	R312	C267258
22.1K(5%)	R105	C43473
AD8232ACPZ-R7	U300	C43216
LDS3985M33R	U100	C222381
SPX1117M3-L-3-3TR	VR200	C6862
4.7uF/6.3V(10%)	C107,C108,C100,C105,C107,C108	C168172
1.4M	R301	C170468
CP2102N-A02-GQFN28	U102	C964632
LESD5D5.0CT1G	D101,D102,D103	C383211
MCP73831T-2ACI/OT	U103	C424093
MPU6050	U400	C24112
3.2K(5%)	R202	C705066
100nF	C200	C1591
4.7uFElectrolyticCap	C201	C2894225
10nF	C101	C1524
0.1uF/50V(10%)	C103,C104,C106,C202	C2655418
19-213/Y2C-CQ2R2L/3T(CY)	D100,D104,D300	C72038
510	R100	C23193
1M	R310,R311,R313	C365401
2.7k	R109	C325419
1.5nF	C306	C1595
47.5K(5%)	R106	C25896
MicroSDCardHolder	CARD200	C370950
10M	R300,R302,R303,R306,R308,R309	C880996
M03PTH	JP300,JP301	C390679
SS-12F44-G5	SW100	C136718
2K(5%)	R107	C881146
22uF/10V(20%)	C102	C108339
470	R102	C23179
360k	R307	C149935

Figure A.6: Bill of Materials for ECG Recorder. Depending on part availability, the total cost of components for one board were approximately £25 at the time of purchase.

Experimental Design and Method

Consent Form

Introduction:

You are being asked to participate in a research study that aims to investigate the effectiveness of an electrocardiography (ECG) in detecting mental stress. The purpose of this study is to better understand how ECG recordings may be used to monitor and detect stress through your ECG signal. Your participation in this study is voluntary, and you may withdraw at any time.

Procedures:

If you agree to participate, you will be asked to wear an ECG recorder device for a designated period of time. The device will be attached to your chest using the chest strap provided and will record your heart's electrical activity during your study periods. You must record your initial start time and stress level. Additionally, they were asked to take note of any changes in stress levels along with the time at which those changes occurred.

Risks and Benefits:

The risks associated with participating in this study are minimal. There is a small risk of discomfort due to the chest strap used to attach the device. The benefits of this study include contributing to research for the potential of gaining a better understanding into the effectiveness of a low-budget ECG device for stress detection.

Confidentiality:

Your privacy and confidentiality will be maintained throughout the study. All data collected during the study will be kept confidential and stored securely. Your data will not be shared with anyone else without your explicit consent. Any personal information collected will be anonymized and kept confidential.

Voluntary Participation:

Your participation in this study is entirely voluntary, and you have the right to withdraw at any time. If you have any questions about the study or wish to withdraw, please contact the researcher.

Consent:

I have read and understand the above information regarding the ECG study. I am unaware of any existing heart conditions or medication that could impact the findings of this study. I understand that my participation is entirely voluntary and that I may withdraw at any. I give my consent to participate in this study and allow the study team to collect and use my data for research purposes.

Participant Signature: _____

Date: _____

ECG Stress Classifier Notebook

The Python Jupyter Notebook for stress classification is publicly available [here](#) and contains the sorting, cleaning, feature extraction, and classification methods used in this study. The structure of the notebook is as following:

- 1. Select Database:** Select either the Spider-Fearful or BrainPatch database.
- 2. Import Libraries and Load Useful Utilities**
- 3. Data Extraction:** Downloads and sorts through databases.
- 4. Preprocessing:** Filtering and signal cleaning.
- 5. Feature Extraction:** R-R peaks, PQRST peaks, EDR, in addition to mean, kurtosis etc.

6. **Feature Selection:** Visualize labelled feature distribution, select desired labels etc.
7. **Traditional Machine Learning Methods:** Linear classification including Naive Bayes, Random Forests, linear/nonlinear SVM.
8. **Convolutional Neural Network:** 1-D convolutional neural network with automatic feature extraction.

Additional BrainPatch ECG Waveforms

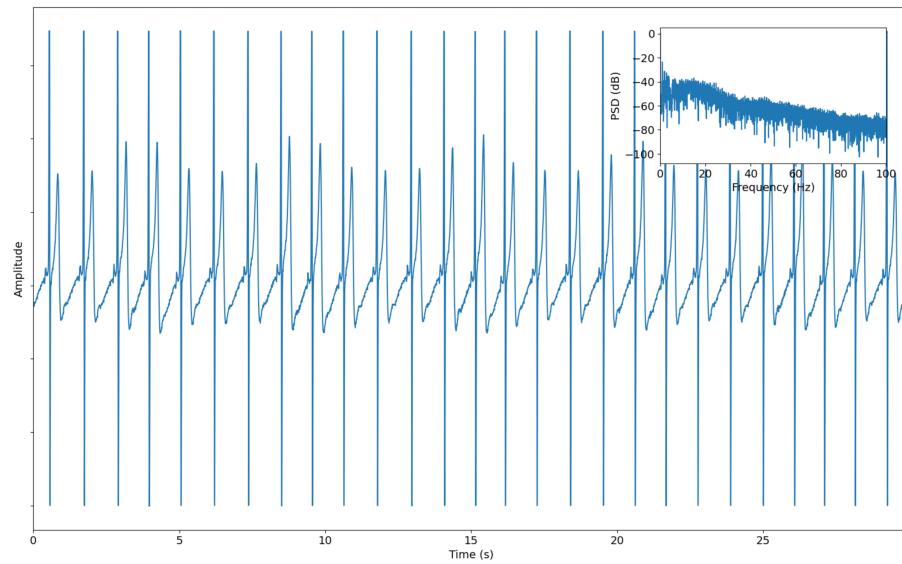


Figure A.7: Raw ECG segment from the BrainPatch device shows minimal baseline drift across a 30 second window.

Additional Spider Fear ECG Waveforms

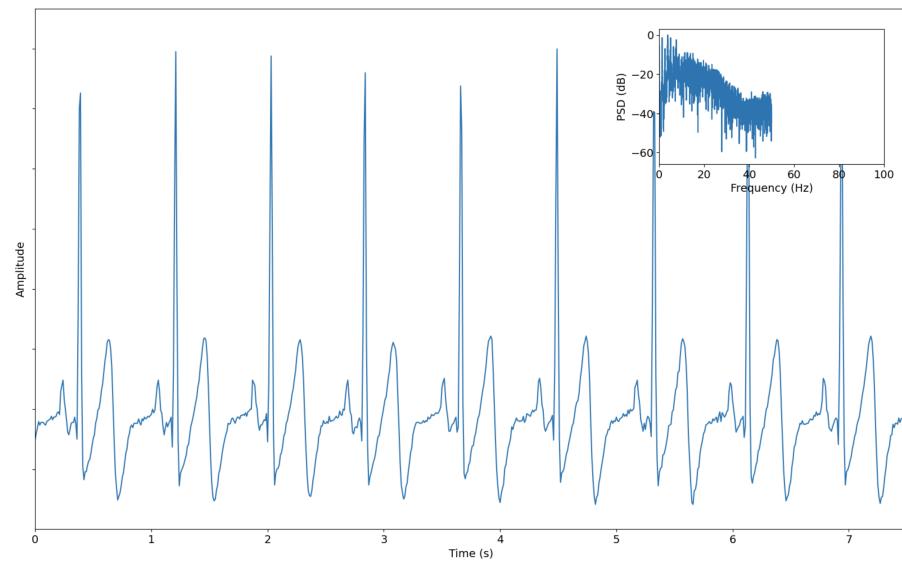


Figure A.8: Raw ECG segment from the Spider Fear device during a participant recording.

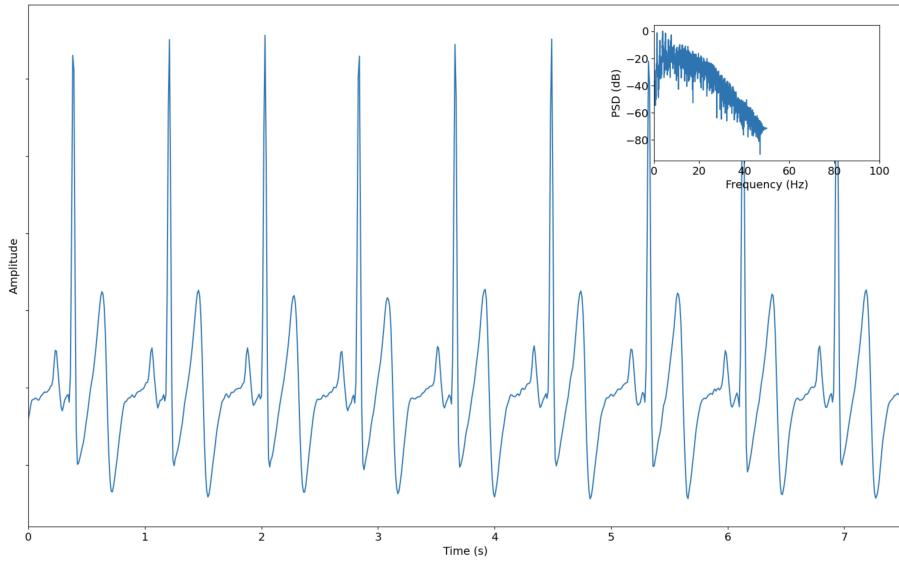


Figure A.9: Cleaned Spider-fear ECG segment after filtering.

Additional Spider Fear Feature Distribution Graphs

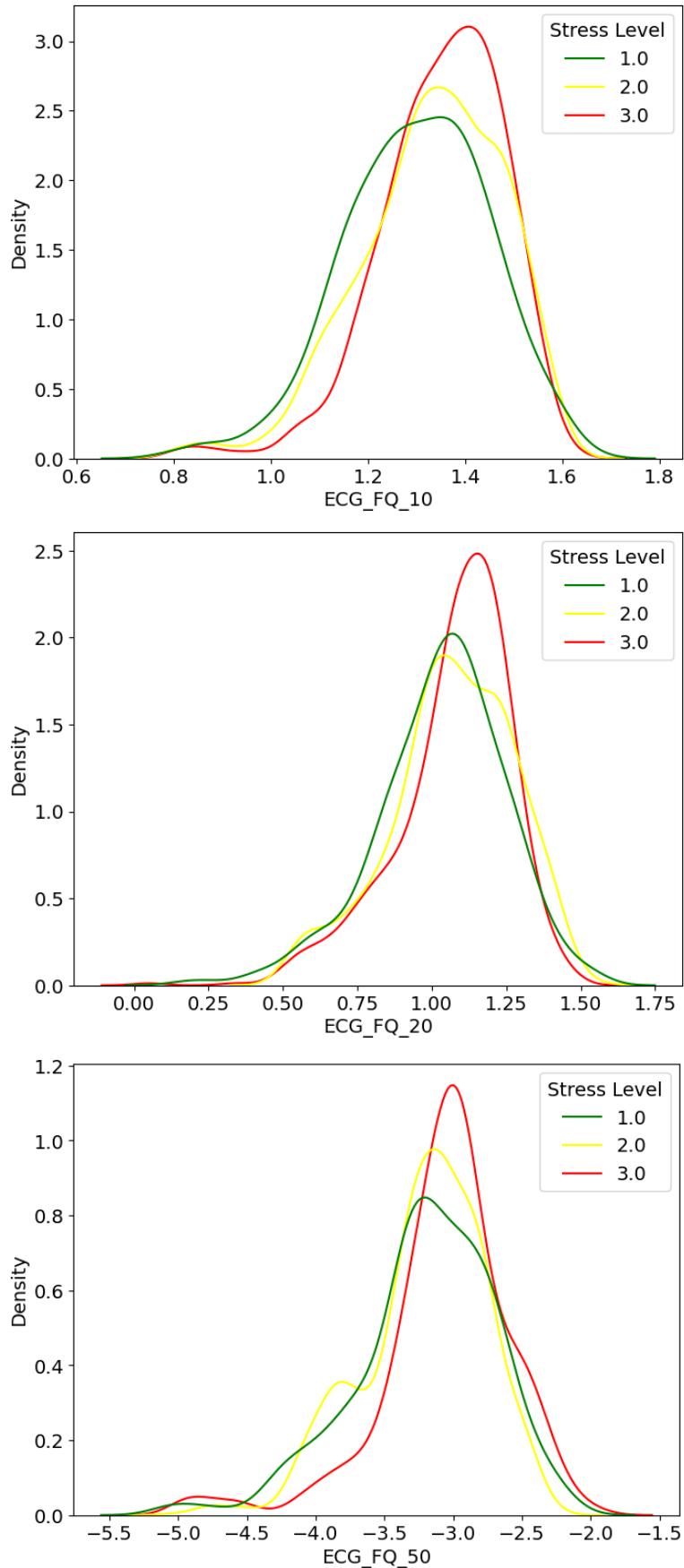


Figure A.10: ECG frequency features for the Spider-fear dataset.

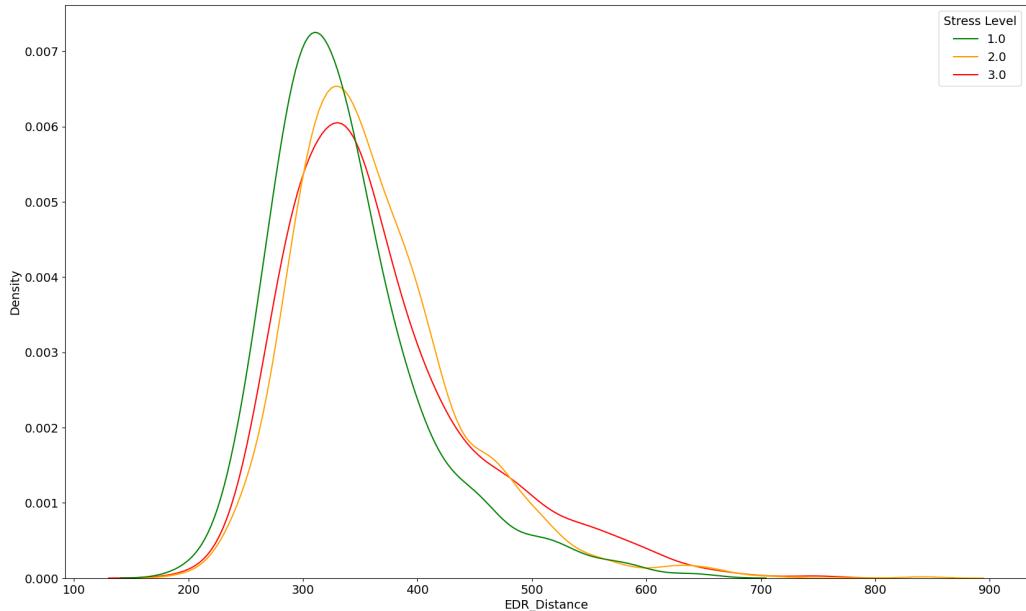


Figure A.11: EDR distance feature for the Spider-fear dataset.

Additional Classifier Results

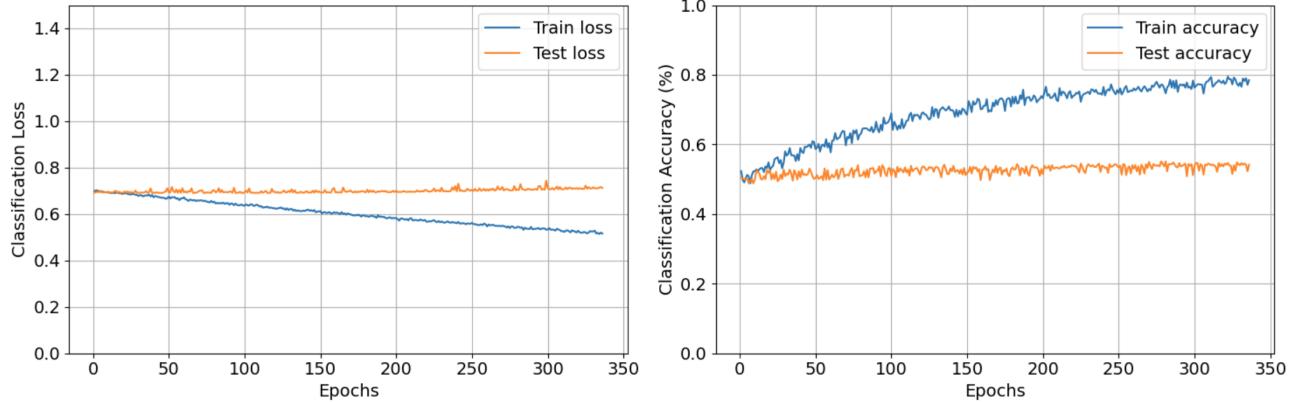


Figure A.12: Learning curve for binary classification using the CNN, showing severe overfitting.

Model	BrainPatch				Spider-fear			
	Binary Classification		Three-level Classification		Binary Classification		Three-level Classification	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
NB	0.667	0.663	0.513	0.511	0.606	0.605	0.454	0.452
SVM	0.814	0.814	0.837	0.833	0.711	0.650	0.650	0.711
RF	0.857	0.857	0.905	0.903	0.872	0.869	0.752	0.749
1D-CNN	0.735	0.730	0.723	0.700	0.893	0.880	0.779	0.773

Table A.1: Model performance comparison for binary and three-level classification using a 30s window for traditional models and 3s window for the CNN.

Model	Window Length							
	5 seconds		10 seconds		20 seconds		30 seconds	
	Binary	Three-level	Binary	Three-level	Binary	Three-level	Binary	Three-level
SVM	0.669	0.468	0.691	0.517	0.721	0.524	0.814	0.837
RF	0.643	0.643	0.682	0.752	0.696	0.774	0.857	0.905

Table A.2: Comparison of different window lengths for Random Forests and SVM.

Model	Sampling Rate (Hz)				
	100	150	250	500	1000
SVM	0.663	0.668	0.741	0.803	0.837
Random Forest	0.882	0.885	0.895	0.903	0.905
1D-CNN	0.679	0.685	0.701	0.729	0.723

Table A.3: The effect of sampling rate on classification accuracy.

Additional Figures

V1-V6 ECG Configuration

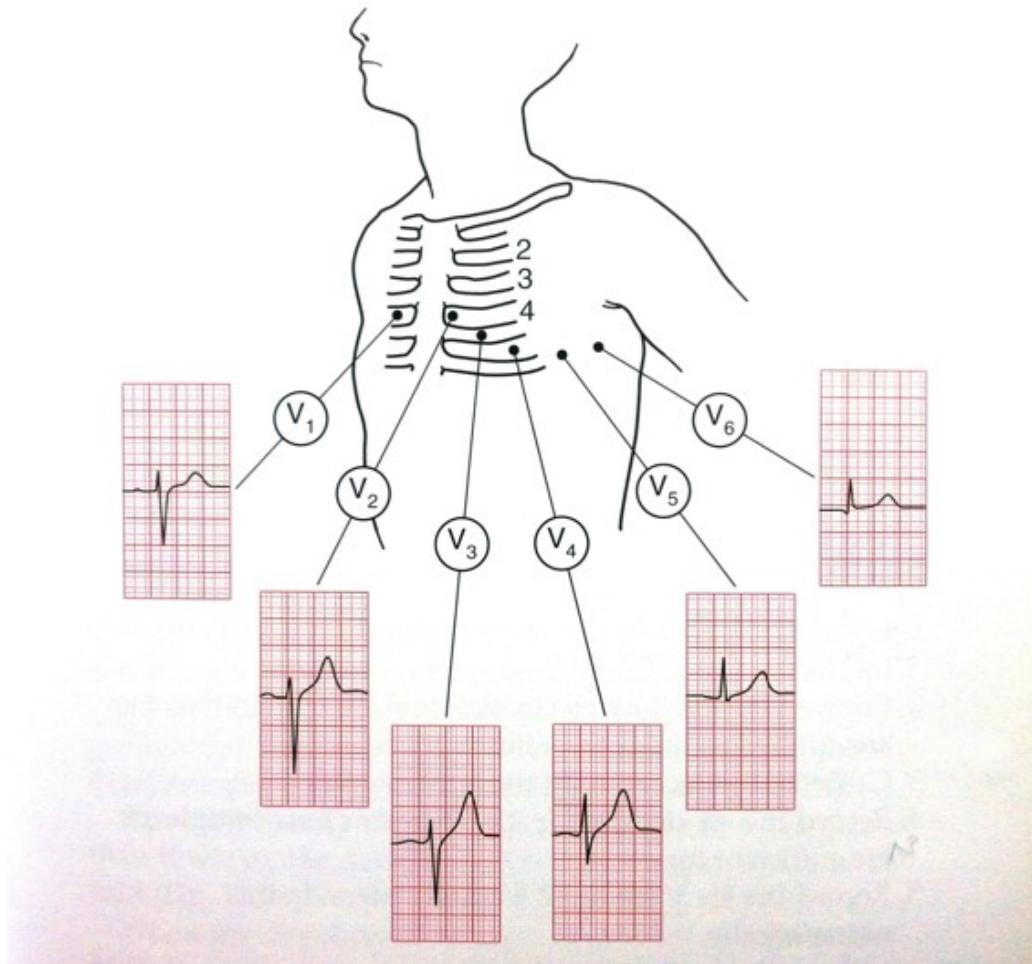


Figure A.13: The electrical activity on an ECG for V1-V6 chest electrodes [111].