

Arrhythmia Classification Using Machine Learning on Cardiac Excitation-Contraction Coupling Data

Syed Shameer Sarwar
Matriculation Number: 20631673

M.Sc. Applied Data Science
Georg-August-Universität Göttingen
syed.sarwar@stud.uni-goettingen.de

March 31st, 2024

Contents

1. Introduction

2. Data Description, Acquisition, and Methodology

- 2.1 Excitation-Contraction Coupling Data Characteristics
- 2.2 Experimental Setup
- 2.3 Raw Data Description
 - 2.3.1 Arrhythmia Events
- 2.4 Available Data Sets
- 2.5 Machine Learning Framework

3. Data Preprocessing

- 3.1 Force
- 3.2 Calcium
- 3.3 Field Potential

4. Ground Truth Generation

- 4.1 Peaks Detection
- 4.2 Window Labeling
- 4.3 Statistical Feature Extraction

5. Modeling and Results

6. Challenges and Future Work

References

1. Introduction

According to statistics from the World Health Organization (WHO), cardiovascular diseases stand as the foremost cause of mortality on a global scale [1]. Conventional approaches to drug development often rely on animal models, yet such methodologies possess inherent limitations in accurately replicating human cardiac physiology. Addressing this disparity, myotwin, an innovative startup headquartered in Göttingen, aims to create the best digital twin of the human heart based on deep learning and generative AI. This cutting-edge technology harnesses artificial heart muscle tissues cultivated from pluripotent stem cells (iPSCs), accurately capturing applicable parameters of cardiac excitation and contraction in response to pharmacological stimuli.

The primary objective of this internship was to discern arrhythmias, denoting irregular cardiac rhythms, within the datasets generated from these artificial cardiac tissues. The clinical management of arrhythmias often necessitates distinct treatment, highlighting the vital importance of a robust detection system. Leveraging a Machine Learning approach, this project attempts to address this challenge. Subsequently, a comprehensive elucidation of the dataset, its characteristics, arrhythmia events, and the strategy for its classification shall be expounded upon in the subsequent sections.

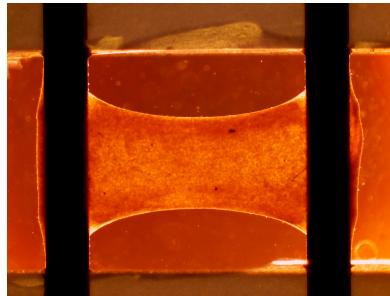


Figure 1: Bioartificial Cardiac Tissues (BCT)

2. Data Description, Acquisition, and Methodology

2.1 Cardiac Excitation-Contraction Coupling Data Characteristics

The excitation-contraction cycle in cardiomyocytes is a vital process that ensures the heart's rhythmic beating. It begins with an electrical signal known as an action potential, which travels through the muscle cell, triggering a cascade of events. This signal prompts the release of calcium ions from internal storage compartments within the cell. Calcium plays a crucial role as a messenger, signaling the cell to contract. As calcium binds to specific proteins within the cell, it induces a change in their shape, initiating muscle contraction. In mammals, this coordinated contraction of cardiomyocytes generates the force necessary to pump blood. In essence, the excitation-contraction cycle orchestrates the synchronized rhythm of the heart, ensuring its efficient functioning and the circulation of oxygen-rich blood to tissues and organs.

2.2 Experimental Setup

The data acquisition system employed in this project consisted of three key components for capturing multi-modal physiological measurements during cardiomyocyte contractions:

- **Multi-electrode Arrays (MEAs)**: These specialized microelectrode arrays facilitated the recording of action potentials. MEAs enable the detection of electrical activity generated by the collective action of cardiomyocytes within the tissue sample.
- **Fluorescent Microscopy**: This imaging technique, in conjunction with a calcium-sensitive fluorescent dye, allowed for the visualization and measurement of calcium transients within the cardiomyocytes.
- **Isometric Sensors**: These sensors were utilized to measure the isometric force generated by the contracting cardiomyocytes. Isometric force measurements assess the contractile strength of the tissue sample without imposing any displacement.

By integrating these three modalities, the data acquisition system provided a comprehensive characterization of the electrical (action potential), calcium dynamics, and mechanical properties (force) of cardiomyocyte contractions, enabling a deeper understanding of their physiological behavior.

2.3 Raw Data Description

The dataset encompasses three distinct parameters, namely the Field potential (mV) delineating electrophysiological activity, followed by Calcium transient intensity (a.u.) and contraction Force (mN). Each experimental trial spans 60 seconds, capturing multiple cycles of contraction, as depicted in Figure 2.a. A representative instance of a single contraction is illustrated in Figure 2.b.

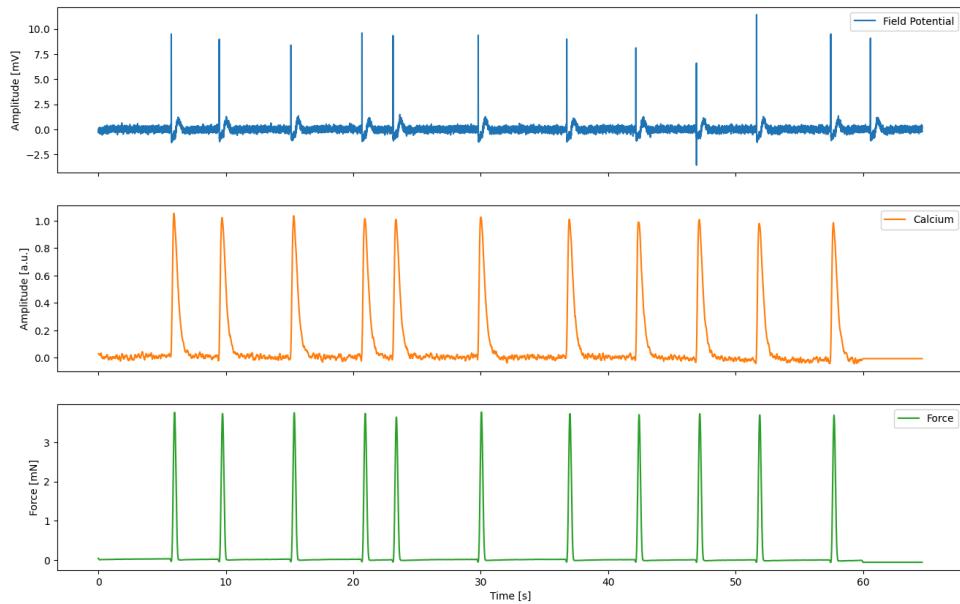


Figure 2.a: Full measurement

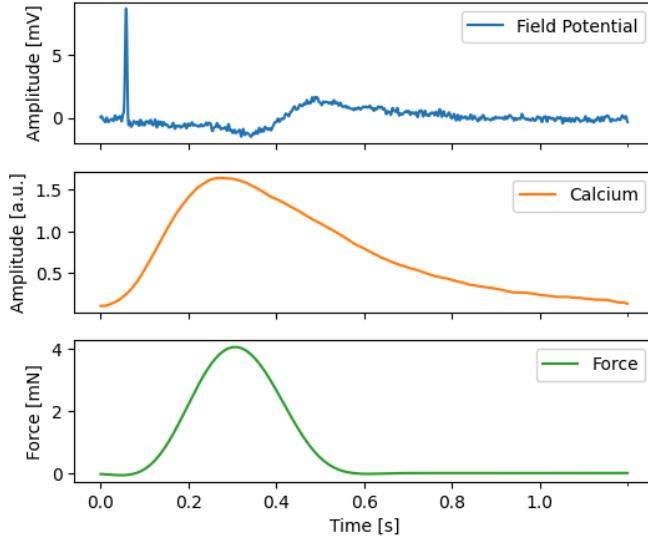
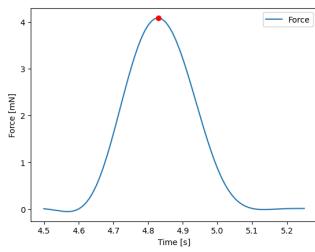


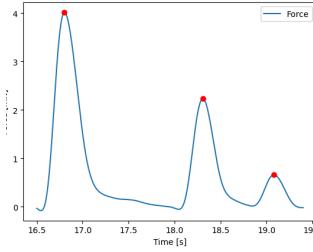
Figure 2.b: Single contraction

2.3.1 Arrhythmia Events

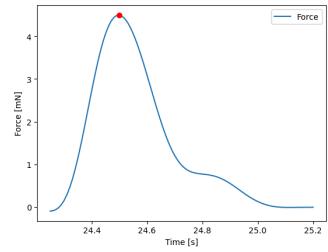
Cardiac arrhythmias, irregular heartbeats, or contractions, exhibit varied presentations across three distinct measurement channels. The manifestation of arrhythmias within these channels is influenced by several factors, including the type and concentration of the administered drug. However, it is reasonable to assume that arrhythmic events occur concurrently across these channels due to their interconnected nature. Notably, in force and calcium measurements, arrhythmias are primarily characterized by premature initiation of a subsequent contraction before the completion of the preceding one. This anomalous behavior may manifest as either closely spaced contractions (Fig 3.ab and 3.bb) or an abnormal decay in the signal waveform (Fig 3.ac and 3.bc). In contrast, identifying arrhythmias within field potential recordings presents a unique challenge. A typical Field potential contraction typically features a sodium peak followed by T-waves (Fig 3.ca). However, discerning arrhythmic events within field potential recordings poses complexities. An arrhythmic contraction may exhibit an additional peak between the sodium peak and t-waves (Fig 3.cb), or it may be obscured by background noise (Fig 3.cc). Furthermore, the identification of t-waves itself is hindered by substantial noise within field potential measurements, further complicating arrhythmia detection within this channel.



[aa] Normal Force



[ab] Arrhythmia: Multiple Peaks



[ac] Arrhythmia: Abnormal Decay

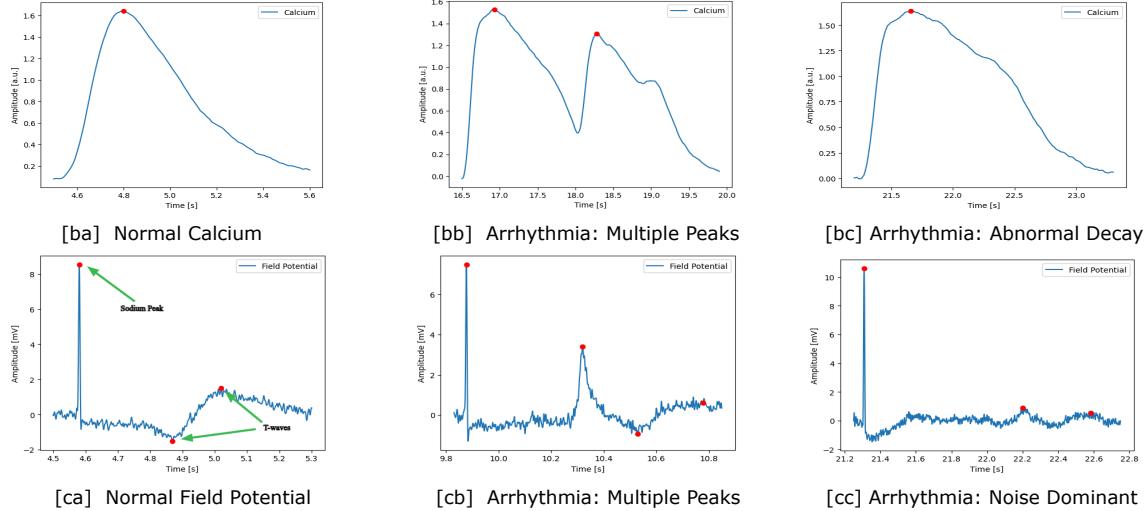


Figure 3: Arrhythmia Forms

2.4 Available Data Sets

The dataset is derived from the application of three distinct pharmacological stimuli: E-4031, Nifedipine, and an experiment called Calcium Titration. These stimuli are administered at varying concentration levels ranging from 3 nanomoles (nM) to 10 micromoles (uM). Notably, the E-4031 drug is known to induce arrhythmia at specific concentration thresholds. The sampling frequency for measurements of force and calcium is set at 100 Hertz (Hz), while for field potential recordings, it is established at 2 Kilohertz (KHz). Furthermore, single channels are utilized for force and calcium measurements, whereas field potential recordings employ 32 channels in the experiment, as depicted in Figure 4.

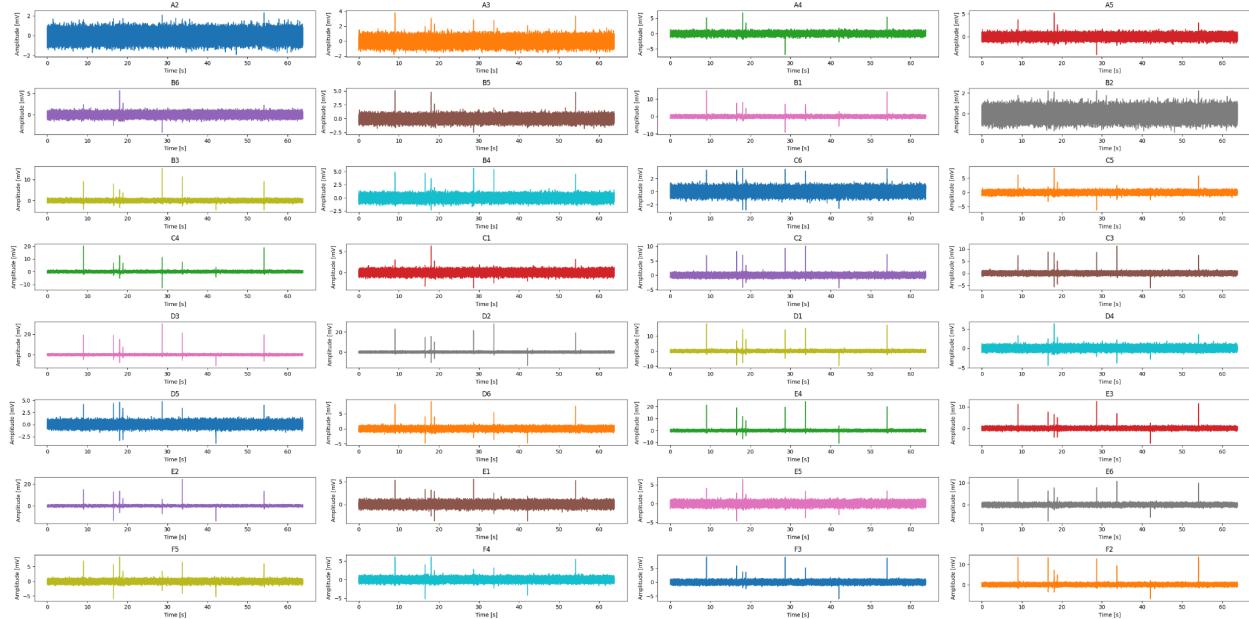


Figure 4: Field Potential (32 channels)

2.5 Machine Learning Framework

We employed a conventional machine learning workflow to address the challenge of arrhythmia classification, which encompasses five primary steps outlined in Figure 5. A comprehensive approach was adopted for data preprocessing, constituting a significant portion of the project timeline. For model selection, we opted for Random Forests due to their resilience and broad utility within the domain. The scientific Python library [SciPy](#) [2] was employed for various signal processing and modeling tasks. Additionally, the Time Series Feature Extraction Library ([TSFEL](#)) [3] was utilized for the extraction of statistical features.

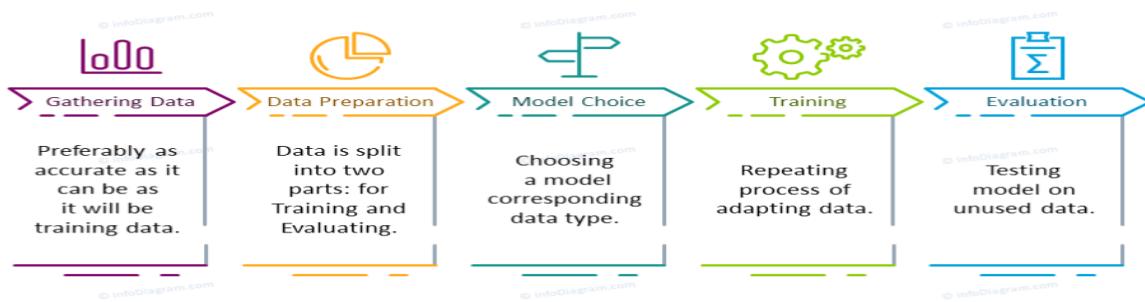


Figure 5: Machine Learning workflow [4]

3. Data Preprocessing

3.1 Force

The preprocessing phase for Force data involves three sequential steps, as depicted in Figure 6. Initially, systematic noise is filtered out, followed by baseline correction and smoothing procedures. To isolate the systematic noise component, a Fourier transform is applied to the Force signal, facilitating the transition of the waveform from the time domain to the frequency domain. Across most measurements, the Fourier transformation reveals a prominent peak around 10Hz (see Fig 7.a), indicative of sinusoidal noise pickup. Given that cardiomyocytes cannot exhibit frequencies exceeding 3.6 Hz or 220 beats per minute [5], measures are taken to eliminate this noise. A 5th-order Butterworth lowpass filter [6], with a cutoff frequency set at 5.2 Hz, attenuates frequencies beyond this threshold while preserving lower frequencies. The frequency response of the 5th-order Butterworth filter at 5.2 Hz, illustrated in Fig 7.b, demonstrates its effectiveness in mitigating noise without affecting much of the critical 3.6 Hz frequency range.



Figure 6: Force preprocessing workflow

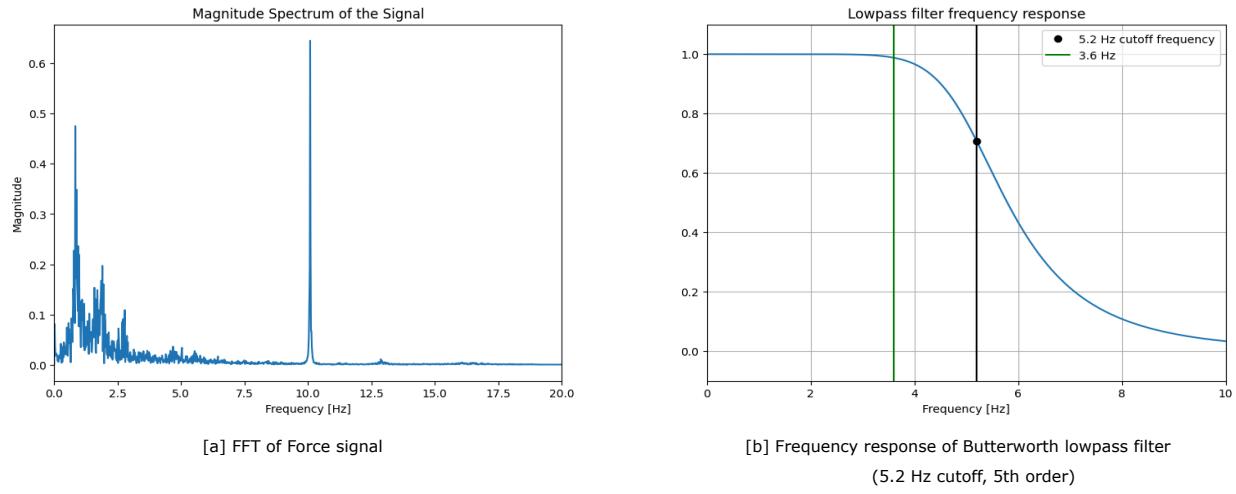


Figure 7: FFT and Butterworth Filter

Following systematic noise removal, a baseline offset correction, also known as pedestal correction, is implemented. The measured signal often exhibits a non-zero baseline offset in force measurements from cardiomyocyte contractions. This offset can be attributed to factors like sensor bias or passive tension in the tissue. Pedestal correction aims to remove this offset by shifting the entire signal to have a zero baseline, facilitating accurate interpretation of the contractile force dynamics. (Fig. 8.a). To achieve this, a histogram of the entire force measurement is constructed. Since noise at the baseline level typically follows a Gaussian distribution, the histogram will exhibit a prominent peak at the left side representing the baseline noise. Conversely, the actual force signal contributions will appear at higher values on the right side of the distribution. (Fig. 8.b). To quantify the baseline offset, a Gaussian curve is fitted around the highest peak corresponding to the baseline noise in the histogram. The mean (μ) of this fitted Gaussian curve is then considered the pedestal value (Fig. 8.c). Subsequently, this pedestal value is subtracted from the entire force measurement, resulting in the final baseline corrected signal (Fig. 8.d).

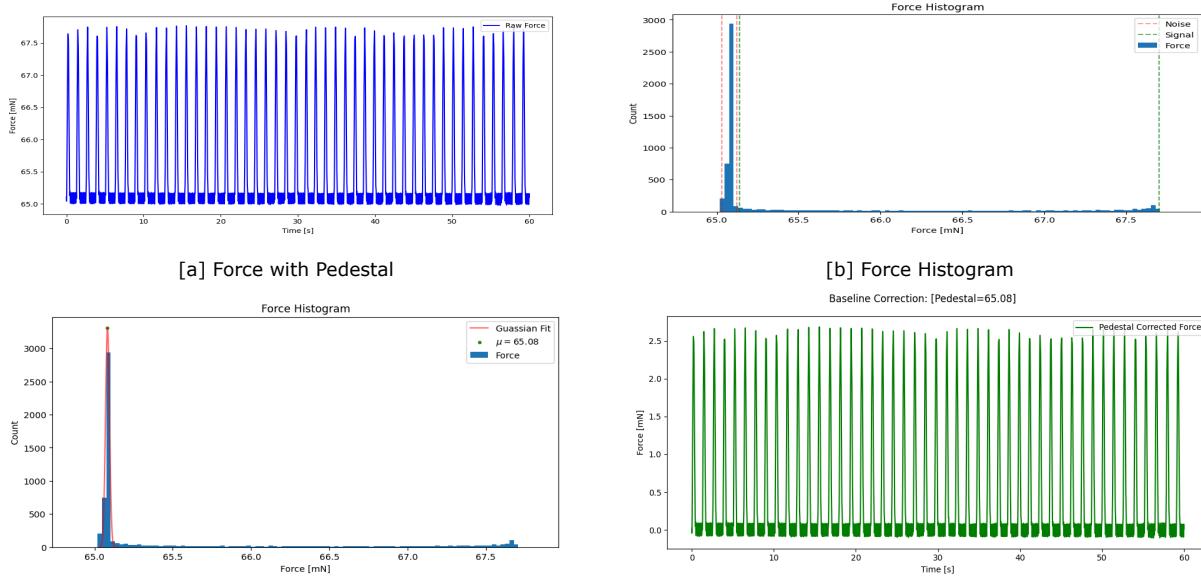


Figure 8: Baseline correction of Force

To further reduce high-frequency noise and enhance the signal-to-noise ratio (SNR) of the baseline-corrected force signal, a [Savitzky-Golay \(SavGol\)](#) [7] filter is applied. The SavGol filter is a polynomial-based smoothing technique that can effectively remove noise while preserving the underlying features of the signal. In this implementation, the SavGol filter utilizes a window size of 21 data points and a second-order polynomial (quadratic).

The final preprocessed force signal obtained after applying all filtering steps can be visualized in Figure 9 for comparison with the raw signal.

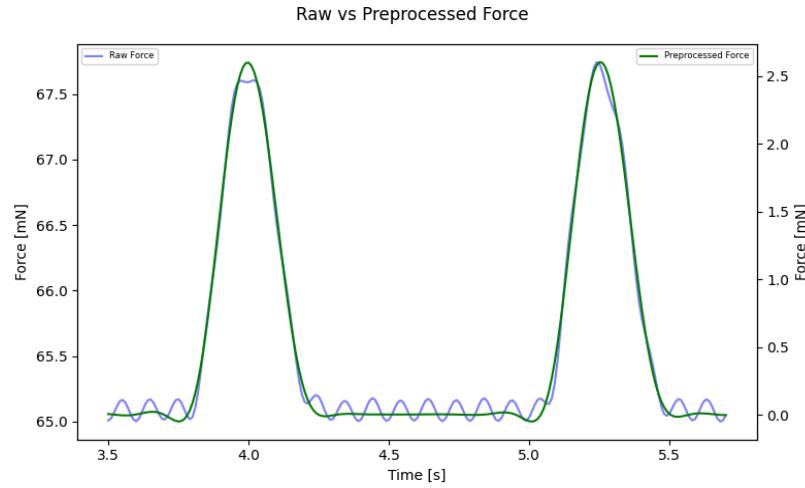


Figure 9: Force Preprocessing

3.2 Calcium

Like the force signal, the calcium signal may exhibit a sloped baseline or pedestal offset. To address this, a two-step baseline correction procedure is implemented. The first step involves estimating the baseline offset using a similar approach to the force signal correction. A Gaussian curve is fitted around the peak corresponding to the baseline noise in the histogram of the calcium signal. The mean (μ) of this fitted Gaussian distribution is considered as the initial estimated baseline value.

Since the calcium signal might contain outliers that can influence the baseline estimation, a more robust method is employed for baseline removal. The Random Sample Consensus (RANSAC) [8] algorithm is utilized to fit a regression line to a subset of data points likely constituting the baseline. This subset is identified by selecting data points within a threshold of ± 2 standard deviations from the mean of the fitted Gaussian distribution in step 1 (Fig. 10a). The RANSAC algorithm is advantageous in this context as it is less susceptible to outliers that may distort the baseline estimation. The fitted regression line obtained from RANSAC is then considered the representative baseline offset. Subsequently, this baseline offset is subtracted from the entire calcium signal to achieve baseline correction.

Following baseline correction, a Savitzky-Golay (SavGol) filter with parameters identical to those used for the force signal is applied to the baseline-corrected calcium signal for noise

reduction and smoothing. Additionally, a time-axis correction is implemented by multiplying the time axis with an experimentally derived factor. This factor accounts for a hardware-induced delay specifically observed in the calcium measurements (Fig. 10b).

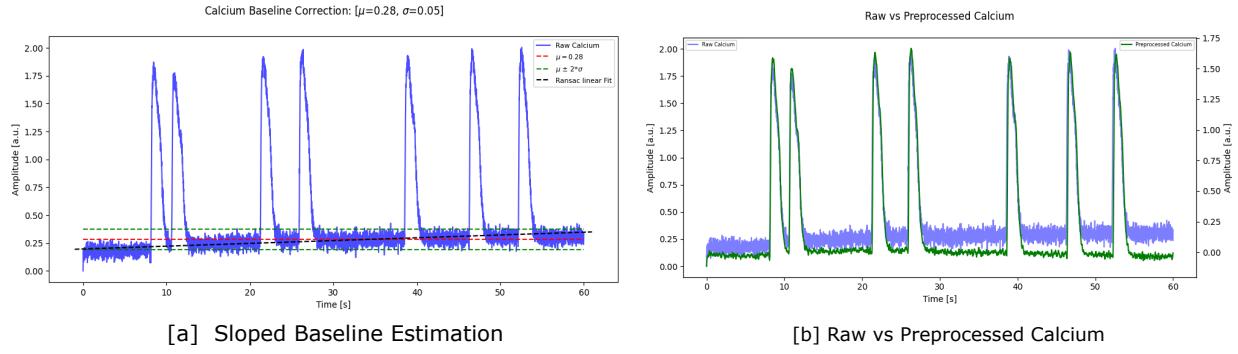


Figure 10: Calcium Baseline correction and Preprocessing

3.3 Field Potential

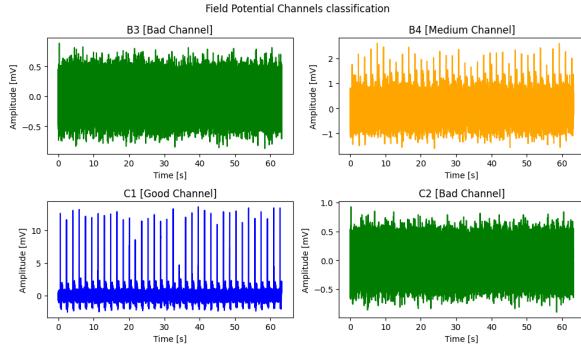
As stated earlier, the field potential signal recordings comprise data from 32 channels. However, these channels exhibit varying signal quality, with some containing substantial noise, others capturing moderate signals, and a subset capturing the full characteristics of the signal, referred to as good channels, (Fig. 11.a). Therefore, a mechanism to classify these channels based on quality was crucial for further analysis.

A supervised learning approach using a Random Forest classifier was implemented to categorize the channels into "bad," "medium," and "good" quality. To train the classifier, a subset of 8 recordings was randomly selected. Within each recording, each of the 32 channels was manually labeled as "bad," "medium," or "good" based on signal quality. Subsequently, statistical features like mean, median, standard deviation, etc., were extracted from each channel using the TSFEL library. These features, along with the corresponding quality labels, were used to train a Random Forest model with 20 estimators and balanced class weights. (A detailed description of the extracted features can be found [here](#). [9])

The trained model was then evaluated on a separate test set containing 45 recordings. This test set included 11 "good" channels, 13 "medium" channels, and 21 "bad" channels. The Random Forest classifier achieved a perfect F1-score of 1.00 on classifying "good" channels, demonstrating its effectiveness in identifying high-quality channels. The classification results are visualized in the confusion matrix presented in Figure 11.b.

Following channel quality classification, only the identified "good" channels were used for further processing. To account for potential signal propagation delays across different sensor locations, the average of these good channels was calculated to obtain a single representative field potential signal (Fig 11.c). This averaging approach enhances timing stability by incorporating signals from various spatial locations.

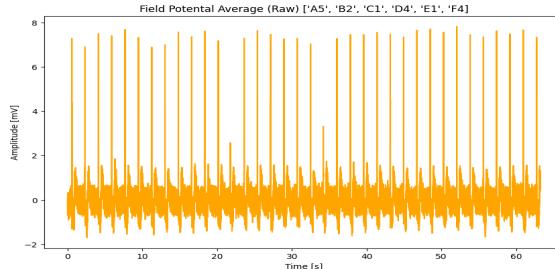
Finally, a 5th-order Butterworth low-pass filter with a cut-off frequency of 180 Hz was applied to the averaged signal (Fig. 11.d). This cut-off frequency was chosen after evaluating a range of frequencies to achieve a balance between noise removal and preserving the essential characteristics of the field potential signal.



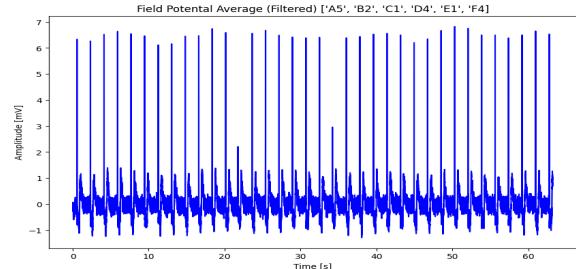
[a] Field potential channel classification

Test Set				
TARGET \ OUTPUT	Bad	Medium	Good	SUM
Bad	20 44.44%	1 2.22%	0 0.00%	21 95.24% 4.76%
Medium	1 2.22%	12 26.67%	0 0.00%	13 92.31% 7.69%
Good	0 0.00%	0 0.00%	11 24.44%	11 100.00% 0.00%
SUM	21 95.24% 4.76%	13 92.31% 7.69%	11 100.00% 0.00%	43 / 45 95.56% 4.44%

[b] Channel classifier confusion matrix



[c] Average Field Potential Signal



[d] Filtered Average Field Potential Signal

Figure 11: Field potential good channels classification and preprocessing

4. Ground Truth Generation

Following the application of preprocessing techniques to each measurement modality (force, calcium, and field potential), we obtain preprocessed signals that represent the underlying physiological events. These preprocessed signals serve as the foundation for the subsequent step of ground truth generation. The process of generating ground truth labels for arrhythmia classification involves a three-step approach, as illustrated in Figure 12. The first step focuses on identifying peaks within the preprocessed measurements. Subsequently, a window of data surrounding each identified peak is selected. These windows are then manually labeled as either "arrhythmia" or "non-arrhythmia" based on the shape of the signal. Finally, similar to the field potential signal processing, statistical features are extracted from the labeled windows using the TSFEL library. It is important to note that, as previously mentioned, arrhythmic events are likely to be reflected simultaneously across all three measurement types (force, calcium, and field potential). Therefore, to improve labeling efficiency, we opted to focus on a single measurement type for peak identification and labeling. Force measurements were chosen due to their superior signal quality and the relative ease of peak identification compared to calcium and field potential signals, which

may be obscured by noise. The following sections will provide a detailed overview of each step involved in the ground truth generation process.



Figure 12: Ground Truth Generation pipeline

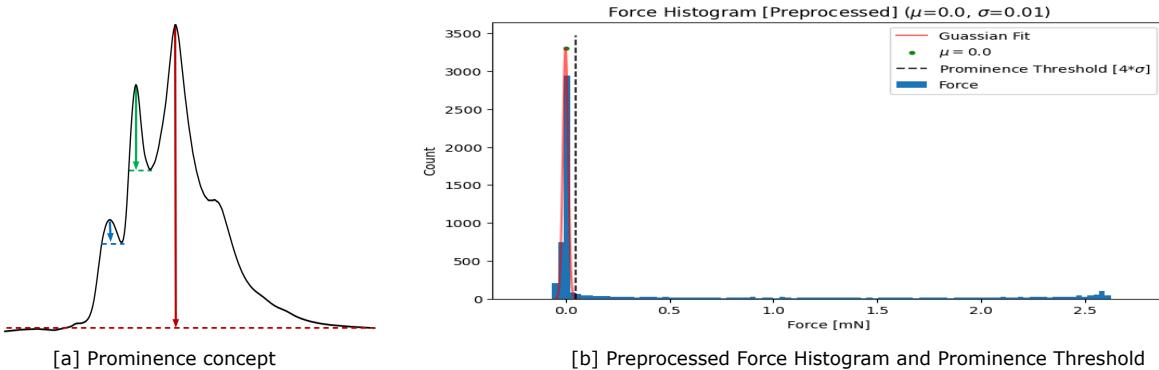
4.1 Peaks Detection

The identification of relevant peaks within the preprocessed force signal is a crucial step in ground truth generation. To achieve this, we employed the “[find_peaks](#)” [10] function from the SciPy library’s signal processing module. However, a dynamic approach to determining the prominence threshold was necessary to effectively distinguish actual peaks from background noise.

The prominence of a peak quantifies its relative significance compared to the surrounding baseline. Mathematically, it represents the vertical difference between the peak’s apex and its lowest contour line (Fig. 13.a). A higher prominence value indicates a more prominent peak, likely to be a genuine indicator of an event.

To determine an appropriate prominence threshold for peak selection, we again leveraged the properties of the preprocessed signal’s histogram. Since the baseline correction step has removed any significant offset, the fitted Gaussian distribution to the histogram peak should ideally have a mean of zero (Fig. 13.b). We then selected a prominence threshold value corresponding to four standard deviations above this mean. This selection is based on the principle that approximately 99.9% of the data points in a Gaussian distribution fall within four standard deviations of the mean. Therefore, peaks with prominence values exceeding this threshold are considered more likely to be relevant and are retained for further analysis.

Figure 13c illustrates the detected peaks within a single force measurement using this prominence-threshold-based approach.



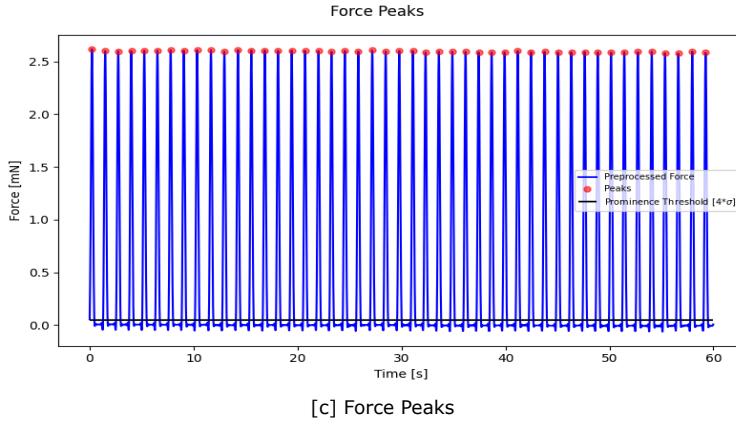


Figure 13: Peaks Detection for Force

4.2 Window Labeling

Following the identification of peaks within the force signal, a window of data surrounding each peak is selected for further analysis. This window encompasses 1.5 seconds on either side of the peak, resulting in a total window size of 3 seconds. Each window is then manually labeled as either "arrhythmia" or "non-arrhythmia". It is important to note that while the peak detection and window selection are performed using the force signal, the same window is subsequently applied to the corresponding time segments in the calcium and field potential signals. This approach leverages the assumption that arrhythmic events are likely to be reflected concurrently across all three measurement modalities. Therefore, the label assigned to the force signal window is directly transferred to the corresponding windows in the calcium and field potential signals.

4.3 Statistical Feature Extraction

Following peak identification and window labeling, a data segmentation strategy is employed to create informative features for arrhythmia classification. Each labeled window is subdivided into multiple sub-windows of varying durations (0.2s to 1.5s) to capture potential signal behavior changes during arrhythmic events. The TSFEL library then extracts 44 statistical features from each sub-window. Importantly, the label assigned during window labeling (arrhythmia or non-arrhythmia) is transferred to the corresponding sub-windows in all measurement types (force, calcium, field potential) assuming synchronized arrhythmic events. This process results in ten distinct feature sets for each measurement type, each associated with a specific sub-window size, enabling exploration of the optimal window duration for arrhythmia classification within each modality. The following provides a visual representation of the entire ground truth generation pipeline.

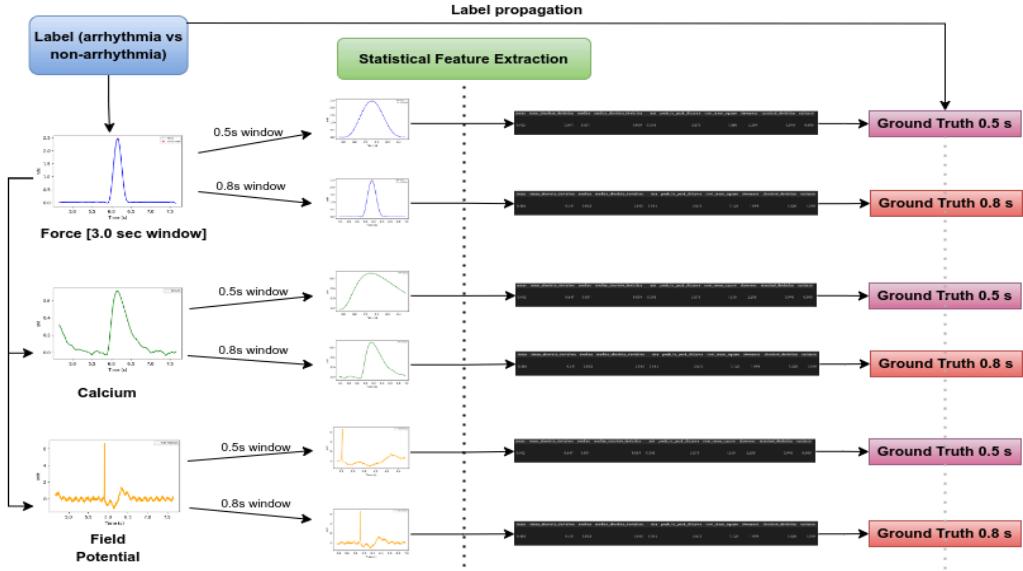


Figure 14: Ground Truth Generation pipeline

5. Modeling and Results

The ground truth generation process yielded a dataset containing 3474 contractions for force and calcium measurements, with 278 labeled as arrhythmic. Some field potential measurements suffered from hardware placement issues and were discarded, resulting in a smaller dataset of 1457 contractions (120 arrhythmic). This imbalanced distribution (only 8% of arrhythmic cases overall) poses a challenge for model training. To address the class imbalance, we employed a Random Forest classifier with balanced class weights and set the number of estimators to 50. As described in Section 4.3, statistical features were extracted from each window size for all measurement types. These features served as the model's input data. The corresponding labels assigned during window labeling in Section 4.2 (either "arrhythmic" or "non-arrhythmic") formed the model's target output. Finally, a stratified train-test split (70% training, 30% testing) was applied to create the training and testing sets used for model development and evaluation.

Individual models were trained for each measurement type (force, calcium, field potential) and various sub-window sizes selected during ground truth generation. The models were evaluated using the F1-score, a suitable metric for imbalanced datasets. Models trained on 1.2-second windows achieved the best test-set F1-score of 0.97 for both force and calcium measurements. For field potential, a 0.3-second window resulted in the highest F1-score of 0.99. Cross-validated mean F1-scores confirmed these findings. Both force and calcium models achieved a mean F1-score of 0.97 across the entire dataset, while the field potential model reached a mean F1-score of 0.98. Figure 16 provides a comprehensive overview of the modeling results on the test dataset. Considering these results, We have successfully developed machine learning models for arrhythmia classification using force, calcium, and field potential measurements, achieving high F1 scores despite class imbalance.

Test set			
TARGET \ OUTPUT	Normal	Arrhythmia	SUM
Normal	960 92.04%	5 0.48%	965 99.48% 0.52%
Arrhythmia	0.00%	78 7.48%	78 100.00% 0.00%
SUM	960 100.00% 0.00%	83 93.98% 6.02%	1038 / 1043 99.52% 0.48%

[aa] Force Model confusion matrix

Class Name	Precision	1-Precision	Recall	1-Recall	f1-score
Normal	0.99	0.01	1.00	0.00	1.00
Arrhythmia	1.00	0.00	0.94	0.06	0.97
Accuracy			1.00		
Misclassification Rate			0.00		
Macro-F1			0.98		
Weighted-F1			1.00		

[ab] Force Model classification scores

Test set			
TARGET \ OUTPUT	Normal	Arrhythmia	SUM
Normal	956 91.75%	2 0.19%	958 99.79% 0.21%
Arrhythmia	3 0.29%	81 7.77%	84 96.43% 3.57%
SUM	959 99.69% 0.31%	83 97.59% 2.41%	1037 / 1042 99.52% 0.48%

[ba] Calcium Model confusion matrix

Class Name	Precision	1-Precision	Recall	1-Recall	f1-score
Normal	1.00	0.00	1.00	0.00	1.00
Arrhythmia	0.96	0.04	0.98	0.02	0.97
Accuracy			1.00		
Misclassification Rate			0.00		
Macro-F1			0.98		
Weighted-F1			1.00		

[bb] Calcium Model classification scores

Test set			
TARGET \ OUTPUT	Normal	Arrhythmia	SUM
Normal	402 91.78%	1 0.23%	403 99.75% 0.25%
Arrhythmia	0 0.00%	35 7.99%	35 100.00% 0.00%
SUM	402 100.00% 0.00%	36 97.22% 2.78%	437 / 438 99.77% 0.23%

[ca] Field Potential Model confusion matrix

Class Name	Precision	1-Precision	Recall	1-Recall	f1-score
Normal	1.00	0.00	1.00	0.00	1.00
Arrhythmia	1.00	0.00	0.97	0.03	0.99
Accuracy			1.00		
Misclassification Rate			0.00		
Macro-F1			0.99		
Weighted-F1			1.00		

[cb] Field Potential Model classification scores

Figure 15: Modeling Results

6. Challenges and Future Work

Our study is limited by the size of the available data, hindering a full evaluation of the model's ability to generalize to unseen data. Additionally, the current approach relies on force measurements for peak detection, introducing a dependency on this modality. Future work will focus on developing robust peak identification methods for noisy calcium and field potential channels, eliminating this dependence. We will also explore methods to improve the robustness of baseline correction and utilize simpler classification algorithms like Decision Trees and Logistic Regression to identify the most informative features from the extracted set. Refining the model by retraining with this subset of features has the potential to improve performance and reduce computational complexity.

References

- [1] <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [2] <https://docs.scipy.org/doc/scipy/>
- [3] <https://tsfel.readthedocs.io/en/latest/>
- [4] <https://www.infodiagram.com/slides/outline-chevron-diagram/>
- [5] <https://upsidestrength.com/blog/maximum-heart-rate/>
- [6] https://en.wikipedia.org/wiki/Butterworth_filter
- [7] https://en.wikipedia.org/wiki/Savitzky%E2%80%93Golay_filter
- [8] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RANSACRegressor.html
- [9] https://github.com/fraunhoferportugal/tsfel/blob/development/tsfel/feature_extraction/features.json/#L292
- [10] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html