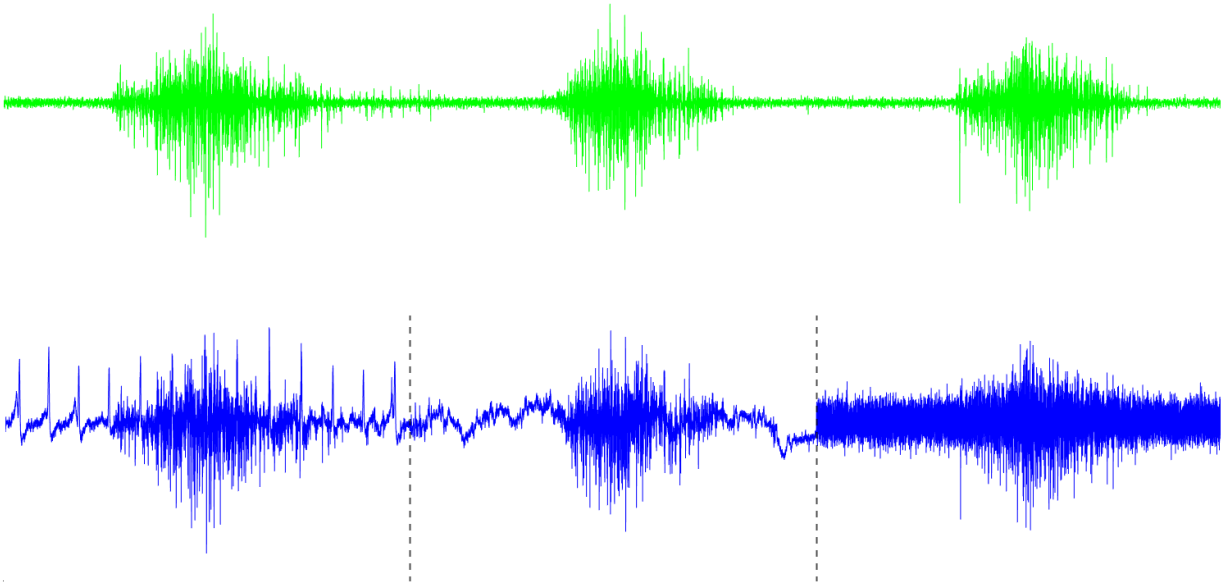
CHALMERS
UNIVERSITY OF TECHNOLOGY



Enhancing sEMG Contaminant Classification Accuracy through Statistical Feature Analysis and Machine Learning

Enabling Improved sEMG Signal Quality

Master's thesis in Biomedical Engineering

ANNIE ABRAHAMSSON EKNEFELT

LUDVIG SWALA

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Enhancing sEMG Contaminant Classification Accuracy through Statistical Feature Analysis and Machine Learning

Enabling Improved sEMG Signal Quality

ANNIE ABRAHAMSSON EKNEFELT
LUDVIG SWALA

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2025

Enhancing sEMG Contaminant Classification Accuracy through Statistical Feature Analysis and Machine Learning
Enabling Improved sEMG Signal Quality
ANNIE ABRAHAMSSON EKNEFELT
LUDVIG SWALA

Supervisor: Morten B. Kristoffersen, Chalmers Industriteknik
Examiner: Giacomo Valle, Chalmers University of Technology

Master's Thesis 2025
Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: MATLAB visualisation of a non-contaminated EMG signal (top) and an EMG signal artificially contaminated (bottom) with ECG interference (first), motion artefacts (second), and white Gaussian noise (third).

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Enhancing sEMG Contaminant Classification Accuracy through Statistical Feature Analysis and Machine Learning
Enabling Improved sEMG Signal Quality
ANNIE ABRAHAMSSON EKNEFELT
LUDVIG SWALA
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Can the most prevalent contaminants in Electromyography (EMG) signals be accurately identified and classified during an EMG examination? EMG with applications are highly affected by signal contaminants, decreasing efficiency and accuracy. The thesis aims in particular to assist home users and non-technical hospital personnel in detecting EMG contaminations, and in the future, guide how to remove contaminants without filtering. This thesis presents two machine learning models for EMG contaminant classification: a single-label model and a multi-label model. A comprehensive feature evaluation was conducted to identify signal features that could differentiate signals of different contaminations and signal-to-noise ratios. Combining time and frequency domain features improved the ability to distinguish between different contaminants at various signal-to-noise ratios, enhancing overall classification accuracy. EMG signals from public databases were artificially contaminated with the three most common EMG contaminants: Electrocardiography (ECG) interference, Motion Artefact (MA) and White Gaussian Noise (WGN). Among several machine learning algorithms, a Random Forest model type achieved the highest accuracy. Two different models, one single-label model and one multi-label model, provide the possibility to either detect the most prominent or all present contaminants. The models utilise five-second segments to classify the EMG signal, allowing for quick feedback on possible contaminations. While offline performance was strong, online validation revealed challenges related to signal variability and generalisation. Still, the approach demonstrates promising potential for classification of EMG signal contaminants in clinical settings, providing the opportunity to improve the quality of EMG signals during examinations or when used with an assistive technology, improving the user's quality of life.

Keywords: Artefacts, Classification, Feature extraction, Machine Learning, Online Validation, Random Forest, Signal Contamination, Signal Feature, Surface Electromyography

Acknowledgements

We would like to thank our supervisor, Morten B. Kristoffersen, for allowing us to work in this fascinating area and for his support, guidance, and expertise throughout the thesis. His encouragement and insights have been invaluable, and we've learned a great deal. We would also like to thank Chalmers Industriteknik and its staff, who supported us with materials and expert input during our thesis work. Their help has been greatly appreciated and encouraged us to further develop our ideas.

Annie Abrahamsson Eknefelt & Ludvig Swala, Gothenburg, May 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AP	Action Potential
CNN	Convolutional Neural Network
DT	Decision Tree
DoF	Degree of Freedom
ECG	Electrocardiography
EMG	Electromyography
ICA	Independent Component Analysis
KNN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MA	Motion Artefact
NB	Naive Bayes
NMJ	Neuromuscular Junction
PSD	Power Spectral Density
PSR	Power Spectrum Range
QC	Quadratic Classifier
RF	Random Forest
s	Seconds
SNR	Signal-to-Noise Ratio
SR	Sarcoplasmic Reticulum
SSC	Slope Sign Changes
SVM	Support Vector Machine
WGN	White Gaussian Noise
ZC	Zero Crossings

Contents

List of Figures

List of Tables

1
Introduction

This chapter provides an overview of the thesis's key objectives. Introduces electromyography (EMG) application areas, myoelectrical control, EMG signal contaminants, and current approaches to classifying contaminants, covering real-time and non-real-time implementations. Finally, the scope of the project is outlined. The project focuses on surface EMG, denoted as EMG.

1.1 EMG Applications

EMG has a wide range of applications, including diagnostics and assistive technology [1, 2]. It can record and stimulate muscle activity and is applied to diagnostics and treatment. Applications can be implemented in medicine, sport, and research.

EMG can be used for diagnostics and assessments of muscle and nerve function, as muscles act as biological amplifiers of nerve signals. As a result, both nerve issues and muscle health issues will be detectable in an EMG [3]. Diseases affecting the nervous system often alter the action potentials (APs) travelling through motor nerves to the muscles, impacting muscle contraction [2]. EMG can diagnose neuromuscular disorders such as myasthenia gravis, corticobasal degeneration, and amyotrophic lateral sclerosis by assessing muscle and nerve function [1, 2]. Additionally, EMG can assess muscle fatigue and perform gait analysis in sports and rehabilitation.

EMG can be utilised in technologies that assist patients in their everyday lives [3, 4]. EMG recordings serve as a method for myoelectric control, enabling muscle signals to control external devices [4]. One common application of myoelectric control is the control of prostheses for individuals with amputated limbs. Exoskeletons are another application of myoelectric control, which can be used for rehabilitation or performance enhancement [3]. In rehabilitation, exoskeletons help patients perform more repetitions, maintain correct posture, and improve overall therapy outcomes, increasing the efficiency of the rehabilitation process. Myoelectric control allows prostheses to be powered rather than passive, enabling more natural movements [4].

1.2 Myoelectric Control

Myoelectric control is a man–machine interface between a device and the wearer [4]. The interface interprets the intended motion from the EMG signal, allowing a device to generate the corresponding mechanical movement [5]. Multiple neural interface options exist, but commercially, muscles are the only viable option for the machine-man interface [4]. In the literature, the nerves and brain are evaluated as options, providing better control at the price of higher invasiveness.

EMG signals are compatible with various myoelectrical control strategies [5]. Different control strategies utilise different action mechanisms. Some are better suited for motion prediction, where future movements are estimated from current or past EMG patterns, and some for motion decoding, where specific motor commands or gestures are inferred from the EMG signals in real time, enabling the control of devices designed to assist or replace movement. All control strategies require good-quality EMG signals.

The degree of freedom (DoF) available affects how many functions a device can provide [4]. Myoelectric control offers greater degrees of freedom than traditional control methods. The DoF is affected by the number of channels available; the more channels, the greater DoF is possible. For myoelectric control, EMG channels are combined with a control method such as thresholds or pattern recognition. The more complex control methods often result in more degrees of freedom, while the less complex ones are more intuitive. Switching methods, used for switching between different functions or different modes, have been included to increase degrees of freedom with the same number of channels, but the switching mechanisms can be unnatural and hinder the control of the device.

For ideal myoelectric control the following criteria must be met [4]:

- Intuitive closed-loop information
- Intuitive control of the device
- Limited computational cost
- Low number of electrodes
- No to little training required
- Robustness to signal variability

EMG signals and myoelectrical control systems have several implementation-related challenges [5]. None of the methods described in the literature fully meet the criteria for an ideal myoelectric device [4]. The performance of myoelectric devices is mainly limited by two factors: crosstalk and variability in EMG signals. Multiple factors influence and introduce variability in the EMG signal, including fatigue, psychological factors, and task performance. Physiological differences between individuals cause variabilities, such as changes in the amplitude and shape of the signal. Crosstalk between different EMG channels limits the DoF that can be accurately extracted from the EMG setup. Contamination of EMG signals poses significant challenges for devices controlled by EMG.

1.3 EMG Contaminants

EMG signals are susceptible to artefacts of numerous origins [1]. EMG signals are complex and have low amplitudes, making their acquisition challenging [3]. Such signals are more susceptible to interference from equipment noise, environmental factors, and activity from the patient's other muscles or physiological systems. The overlap between the EMG power spectrum and contaminant frequencies creates a challenging signal processing task, as significant information may be lost [1].

Crosstalk from other muscles, electrode movement, impedance changes, amplifier saturation, power line interference, electrocardiography (ECG), White Gaussian noise (WGN) and poor electrode contact resulting in motion artefacts (MA), among other sources, create artefacts that tend to contaminate the EMG signal [1, 6]. The effects of ECG, MA, and WGN contaminants on EMG signal are presented in Figure 2.1, in Chapter 2.3, all contaminants present lower the quality of EMG signals. Different contaminants have different characteristics and require specific filters to be suppressed in the signal processing phase [4]. As described in Chapter 1.2, it is common to base control algorithms on EMG signals. For the control algorithm to perform as expected, the signal-to-noise ratio (SNR) must be as high as possible. To achieve a high SNR, filtering is often necessary, and selecting the appropriate filter ideally requires knowledge of the type of contaminant present. Due to the overlap in frequency components between contaminants and the EMG information of interest, filtering contaminants while preserving muscle information to increase SNR is challenging.

1.4 Current Solutions

Previous work has investigated how to classify contaminants in EMG signals. The classification challenge is divided into two parts: methods for data processing and methods for classification. For all previous work, classification accuracy is highly dependent on the SNR; a low SNR resulted in a high classification accuracy, while a higher SNR decreased the accuracy [7].

The information used for classification can be generated from different processes. Information can be gathered from the time and/or frequency domain. The literature identifies three distinct data input strategies for classifying signal contaminants: extraction of signal features, use of time-series signals, and analysis of signals separated by Independent Component Analysis (ICA). Feature extraction can be performed in the time domain, frequency domain, or time-frequency domain, enabling a range of classification approaches. Table 1.1 summarises the data input strategies for EMG signal contaminant classification.

Table 1.1: EMG signal representation methods used for classification of contaminants from literature review

Method	References
Signal Features	[7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]
Signal-ICA Components	[26]
Time-Series Signal	[8], [27], [28]

The features extracted from EMG signals are computed in either the time or frequency domain. Time domain features are presented in Table A.1 in Appendix 1, and frequency domain features are provided in Table A.2 in Appendix 1.

Multiple methods can identify contaminants within an EMG signal [6]. Machine learning methods are gaining popularity, with a wide range of possible applications [7, 8, 9]. The data used for classification typically consists of signal features, time-series signals, or signal-ICA components, as outlined in Table 1.1. The classification methods identified for EMG are summarised in Table 1.2.

Table 1.2: EMG signal contaminant classification methods from literature review

Method	References
Actor-Critic Reinforcement Learning	[16], [17]
Convolutional Neural Network	[8], [9], [28]
Decision Tree	[8]
Fuzzy Inference System	[12]
K-Nearest Neighbour	[8]
Linear Discriminant Analysis	[9], [18]
Long Short-Term Memory	[8], [9]
Naive Bayes	[8]
Probability Density Function Profile	[13]
Quadratic Classifier	[9]
Recurrent Neural Network	[7], [27]
Self-Organising Map	[20]
Support Vector Machine	[8], [10], [11], [14], [15]

A strong correlation between SNR levels and challenges in classification accuracy is observed; higher SNR corresponds to reduced signal contamination, which increases the complexity of contaminant classification and decreases accuracy [8, 9, 11, 12, 13, 14, 15, 17, 19, 28]. Studies such as [11, 13, 28] report a rapid decline in accuracy for SNR values exceeding 0 dB, whereas [14, 17] indicate a similar trend for SNR values greater than -10 dB. To ensure a robust user system, classification must be

reliable at SNR levels high enough for the EMG signal to be considered clean, or sufficiently clean to allow the system to function accurately. The SNR threshold that differentiates between clean and contaminated EMG signals is provided in Table 1.3.

Table 1.3: SNR thresholds for defining clean EMG Signals identified in the literature

SNR [dB]	References
15	[8]
18	[9], [11], [14]
20	[10]

A common theme for all previous work is the rapidly decreasing classification accuracy with increasing SNR. Table 1.4 provides an example of the decrease in accuracy with the increase in SNR for Long Short-Term Memory (LSTM) models. The accuracy presented in the table results from the combinations of parameters that achieve the highest accuracy per SNR level in [7].

Table 1.4: Highest achieved classification accuracy for LSTM model per SNR level, reported in [7]

SNR [dB]	Accuracy
-40	0.9875 ± 0.0948
-30	0.9960 ± 0.0339
-20	0.9632 ± 0.1649
-10	0.6845 ± 0.4181
0	0.2717 ± 0.4252
10	0.2024 ± 0.3972
20	0.2002 ± 0.3961
30	0.2002 ± 0.3961
40	0.2002 ± 0.3961

Hardly any previous work has applied the classification of contaminants in practical situations, enabling classification during EMG examination [9, 18]. Real-time implementations are affected by the correlation between SNR levels and classification accuracy. One article demonstrates the feasibility of real-time implementation by utilising a Quadratic Classifier (QC), Convolutional Neural Network (CNN), LSTM, or Linear Discriminant Analysis (LDA) on an Arduino Mega 2560 [9]. The article defines SNR levels above 18 dB as uncontaminated, but does not provide additional details on the distribution of SNR levels within the training and testing datasets. Another method used a laptop PC for real-time implementation, achieving a near-zero delay for classification [18]. The article provided no information on the SNR levels defined as contaminated or non-contaminated, or on the distribution of the SNR levels within the training and testing datasets.

1.5 Project Scope

The purpose of this thesis is to develop a model that offers information on potential EMG contaminations. The thesis aims in particular to assist home users and non-technical hospital personnel in detecting EMG contaminations, and in the future, guide how to remove contaminants without filtering. This could assist in improving EMG signal quality, since removing EMG contamination without removing EMG information. By improving quality, various applications, such as assistive technologies and diagnostic tools, can operate more effectively and improve users' quality of life.

The project's goal is to answer the research questions formulated below:

1. What are the most prevalent contaminants in EMG?
2. Which signal features provide the most reliable basis for EMG signal contaminant classification?
3. Which machine learning model delivers the highest accuracy for EMG signal contaminant classification?
4. Can EMG contaminants be classified while using EMG?

Some limitations must be accounted for in the scope of the project. The exact number of possible contaminants in EMG signals is unknown. However, this project will focus on the three most commonly reported contaminants in the literature. Within the scope, only one data input strategy and one classification method will be selected. The project aims only to include data from intact subjects.

2

Theory

This chapter outlines the key theories behind skeletal muscle function, EMG, EMG signal contaminants, signal features, and the machine learning algorithms employed for their classification.

2.1 Skeletal Muscle Function

Skeletal muscle is the most common of the three types of muscle within the human body. Approximately 40% of the total body weight is composed of the skeletal muscles [29]. The skeletal muscle has many purposes; among them stabilising joints, producing movements, maintaining body temperature, and sustaining body posture [30]. The majority of the skeletal muscles are attached to the skeleton via tendons. By contracting the muscle, the tendons pull the skeleton, creating a movement. Combinations of multiple muscles create more advanced movements. All physiological movements are produced by muscle contractions, from everyday movements to those of elite athletes. The level of contraction decides the amount of force created by the muscle. The greater the contraction, the greater the force generated. The more contractions a muscle performs and the greater the load it experiences over time, the stronger it becomes and the more force it can generate. Skeletal muscle needs to be able to adapt to rapidly changing functional demands.

Skeletal muscle tissue is organised into multiple structural levels, each formed by bundling smaller units together [30, 31]. At the highest level, muscles consist of fascicles, bundles of muscle fibres (myofibres). The overall size of a muscle primarily depends on the number of muscle fibres it contains [29]. These fibres are classified into three types: Type I, Type IIa, and Type IIb, each with distinct characteristics that influence their recruitment for different activities [30]. A summary of these fibre types and their properties is provided in Table 2.1. Each myofibre consists of numerous myofibrils, which in turn contain two main types of protein filaments: actin (thin filaments) and myosin (thick filaments). These filaments are arranged in repeating structural units called sarcomeres, which give skeletal muscle its striated appearance. Sarcomeres are the smallest functional units of muscle tissue and play a crucial role in muscle contraction.

Table 2.1: Skeletal muscle fibre types and physiological characteristics

Fibre Type	I	IIa	IIb
Fatigue Rate	Slow	Intermediate	Fast
Myoglobin Concentration	High	High	Low
Mitochondria Concentration	High	High	Low
Twitching Rate	Slow	Fast	Fast
Example Activity	Keeping posture	Walking	Sprinting

Skeletal muscles allow for the creation of body movements by converting chemical energy into mechanical energy [29]. Mechanical energy is presented as the movement of joints, a result of the contraction of muscles. A voluntary movement starts in the motor cortex of the brain. Information, the impulses, are transported through the efferent pathways to the skeletal muscles [31]. The impulse consists of APs, rapidly changing electrical signals. The AP travels along motor nerves towards the connection between the motor nerve and muscle fibres, the neuromuscular junction (NMJ) [31, 32]. The change in potential causes the motor nerve to secrete acetylcholine (ACh) into the NMJ. The ACh affects the muscle fibres locally to open ACh-gated cation channels. This allows large quantities of sodium (Na) ions to diffuse past the membrane. This causes a local depolarisation, which opens voltage-gated Na-channels, and this initiates an AP at the membrane. The AP's depolarisation of the membrane causes the sarcoplasmic reticulum (SR) to release large quantities of calcium (Ca) ions stored within the reticulum. The actin and myosin within the muscle fibre experience an attractive effect from the Ca-ions, causing them to slide alongside each other, contracting the muscle. Once the calcium is released from the SR, the Ca-membrane pumps pump back Ca-ions into the SR. This removal of Ca-ions from myofibrils causes the muscle contraction to ease until the next AP occurs. The more Ca-ions, the stronger and longer the contraction, and more Ca-ions are released as a result of multiple APs.

2.2 Electromyography

The AP generated over the efferent nerve and muscle fibre membranes can be measured [1, 4]. The AP over muscle fibres are directly associated with the AP over the nerve and corresponds to the level of muscle contraction. One AP over the nerve gives rise to multiple APs over the whole muscle, allowing the muscle to work as a natural amplifier for measuring muscle activation. Measuring APs in the muscle is denoted as EMG and can be performed invasively or non-invasive. The invasive method includes inserting electrodes, in the shape of needles or hooks, into the muscle [1]. The non-invasive method utilises surface electrodes placed on the skin above the muscle, either dry or wet. The intramuscular EMG are more precise, measuring the local area and the surface based method measures AP from a wider area beneath the electrode. Since the neural information is the same for surface and intramuscular EMG, surface EMG is often preferred [4]. Surface EMG typically has a power spectrum between 2 Hz and 500 Hz, while intramuscular EMG ranges from

10 Hz to 10,000 Hz [1]. The measured signal is the total signal from all APs over the muscle fibre membranes within the recording area.

EMG consists of electrodes, amplifiers and filters, an EMG acquisition instrument, an analogue-to-digital converter, a power supply, a data transfer solution and a host computer that, among other things, presents the final signal to the handler [3].

2.3 Contaminants

ECG artefacts occur when EMG electrodes pick up electrical activity from the heart, which is especially prevalent when EMG is measured in areas close to the chest, neck, back or diaphragm [6]. These artefacts manifest as sudden spikes in the EMG signal, corresponding to the heart's characteristic QRS complex, and are primarily concentrated in the 10 Hz to 20 Hz frequency range.

MA in EMG signals arise due to the disruption of the half-cell potential at the electrode-gel interface and skin deformation, leading to variations in skin potential [6]. These artefacts are highly variable and predominantly occur at low frequencies (<20 Hz). In the EMG signal, MA typically appear as slow baseline drift and irregular, sharp transients or spikes that do not correspond to muscle activity. They can significantly distort the EMG signal, making it challenging to extract meaningful information.

Random fluctuations in the measurement system often introduce WGN in EMG recordings, including electrical components such as amplifiers and filters [6]. WGN is characterised by its broad frequency spectrum and is generally present across all frequencies. It appears in the signal as seemingly random variations. This can be particularly problematic for EMG due to the low amplitude of the signal.

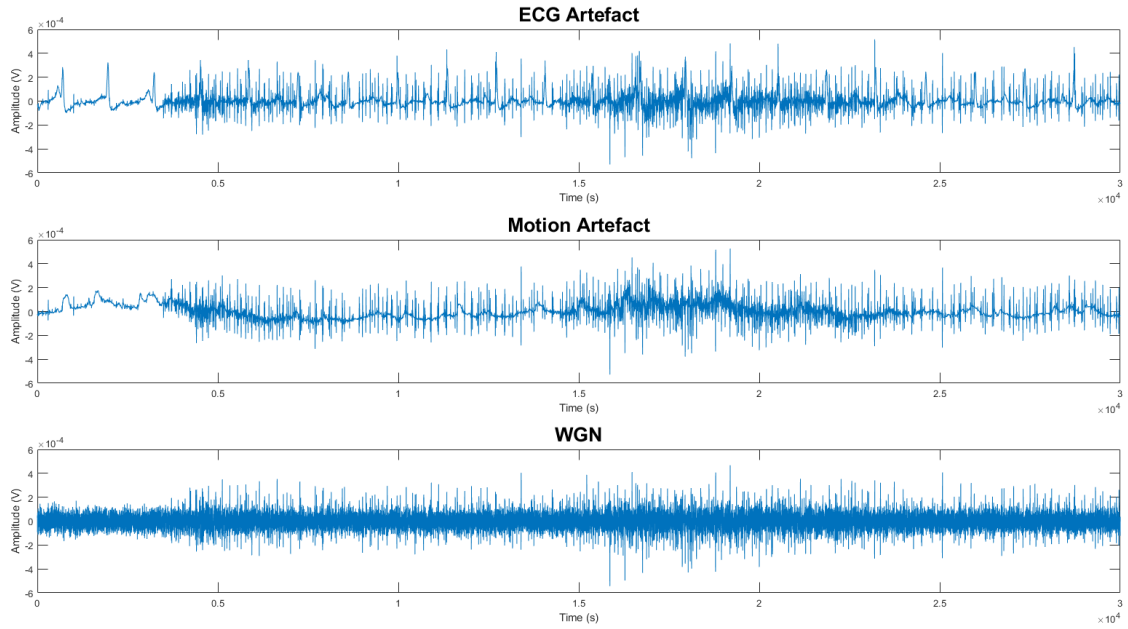


Figure 2.1: EMG signal contaminated with ECG, MA, and WGN artefacts, at 0 dB SNR

2.4 Features

Signal features are numerical descriptors that help characterise important patterns in a signal. The features extract meaningful information from raw EMG data, facilitating the identification and classification of contaminants in the signal. This section presents the subset of signal features selected for final use after the feature evaluation process described in Chapter 3.4. Although a larger number of features were initially considered, only those showing robust performance across varying contamination types and SNR levels were retained. The following features were therefore identified as the most discriminative and are described in detail below.

The number of slope sign changes (SSC) measures the number of times the second derivative of the signal changes sign within a window [25, 33]. SSC is calculated as:

$$SSC = \sum_{i=1}^{N-2} \mathbf{1}_{\left(\frac{d^2 x_i}{dt^2} \neq 0\right)}$$

Where x_i is the EMG signal at time step i, N is the length of the signal, and $\mathbf{1}$ is the indicator function that returns 1 if the second derivative changes sign.

Zero crossings (ZC) are calculated by counting the number of times the signal crosses the zero axis within a window [25, 33]. The ZC count is given by:

$$ZC = \sum_{i=1}^{N-1} \mathbf{1}_{(\text{sign}(x_i) \neq \text{sign}(x_{i+1}))}$$

Where N is the length of the signal and $\text{sign}(x_i)$ is the sign function that returns +1 for positive values and -1 for negative values.

The Hurst exponent, H, measures the self-similarity or long-term memory of the time series [34]. The Hurst exponent is calculated using a rescaled range (R/S) analysis method, which estimates how the range of a signal grows with its length over different time windows:

$$H = \frac{\log\left(\frac{R}{S}\right)}{\log(N)}$$

Where R is the range of the cumulative sum of the mean-adjusted signal, S is the standard deviation of the signal, and N is the length of the signal.

The Power Spectral Density (PSD) Deformation Ratio measures the relative change in the spectral characteristics of the signal. It is sensitive to changes in symmetry and peaking of the power spectrum [25]. It is calculated as:

$$\text{PSD Deformation Ratio} = \frac{\sqrt{\frac{M_2}{M_0}}}{\frac{M_1}{M_0}}$$

Where M_0 is the total power, M_1 is the first moment (mean value), and M_2 is the second moment (variance) of the PSD.

PSD Deformation is a custom spectral feature that reflects deviation from spectral balance. It is defined as:

$$\text{PSD Deformation} = \frac{M_2}{M_0} - \frac{M_1}{M_0}$$

This feature is similar in structure to the variance of central frequency, as described in [35], but differs in that it subtracts the linear mean rather than its square.

L-Kurtosis is a measure of the sharpness of the signal's probability distribution [13]. In this project, it is calculated as:

$$\text{L-kurtosis} = 0.5\left(3 - \frac{Q_{75} - Q_{25}}{Q_{95} - Q_5}\right)$$

Where Q_5, Q_{25}, Q_{75}, and Q_{95} are the quantiles of the signal at the 5th, 25th, 75th, and 95th percentiles, respectively. The subtraction of 3 normalises the value so that a normal distribution (which has a kurtosis of 3) results in an L-kurtosis near 0 [12]. The 0.5 scaling factor reduces the impact of outliers and ensures the measure reflects the distribution's sharpness within a manageable range. Positive values indicate sharper peaks, while negative values indicate flatter distributions.

The Power Spectrum Range (PSR) is a custom feature introduced in this project, inspired by the Drop in Power-density Ratio [25]. It quantifies the relative difference between the maximum and minimum values of the PSD, normalised by the maximum value. It is calculated as:

$$\text{PSR} = \frac{\max(\text{PSD}) - \min(\text{PSD})}{\max(\text{PSD})}$$

This feature captures the dynamic range of spectral energy and is particularly sensitive to abrupt spectral changes, such as those caused by transient events or signal disturbances.

2.5 Machine Learning Algorithms

This section outlines the machine learning algorithms for EMG contaminant classification, focusing on supervised learning methods, including traditional classifiers, probabilistic models and ensemble techniques.

Support Vector Machine (SVM) is a supervised machine learning algorithm commonly used for classification tasks [36]. SVM works by finding the optimal line, or hyperplane if in a higher dimension, that maximises the distance between the closest data points of different classes. For linearly separable data, SVMs efficiently classify new data based on which side of the hyperplane they fall on. However, much real-world data is not linearly separable. In such cases, nonlinear SVMs can be used as they leverage kernel functions to transform the data into a higher-dimensional feature space where it becomes linearly separable. This increases the algorithm's complexity, which in turn increases the computational load and the risk of overfitting.

K-Nearest Neighbour (KNN) is a supervised machine learning algorithm used for classification and regression [37]. For classification, KNN assigns a class label to a new data point based on the labels of its K-number of closest neighbours. If multiple classes are present among the K neighbours, a majority vote takes place, where the new data point is assigned the most frequent class among its neighbours. Some variations of KNN may also weigh votes based on the distance of each neighbour, giving closer points more influence in the classification.

Linear Discriminant Analysis (LDA) is a supervised machine learning algorithm used for classification [38]. LDA aims to find a projection that maximises the separation between the means of different classes while minimising the variance within each class. LDA assumes that all classes share a common covariance matrix, which results in linear decision boundaries. It identifies a linear combination of features that best separates two or more classes. This is typically achieved by projecting data from multiple dimensions into a lower-dimensional space (often one dimension for binary classification), simplifying the classification task.

Quadratic Classifier (QC), also known as Quadratic Discriminant Analysis, works very similarly to LDA [39]. It differs by allowing each class to have a unique covariance matrix. This flexibility leads to quadratic decision boundaries, enabling the QC algorithm to model more complex class distributions.

Decision Tree (DT) is a supervised machine learning algorithm that can be used for either regression or classification [40]. DTs have a hierarchical, tree-like structure consisting of a root node, branches, internal nodes and leaf nodes. The tree is built by recursively splitting the dataset into subsets based on the feature values, aiming to achieve the best possible separation between features. At each internal node, a decision is made based on a feature threshold, with each branch representing the possible outcomes of that decision. The leaf nodes at the end of the branches represent the final predicted class (for classification). While DTs are powerful and interpretable, a complex DT tends to overfit and does not generalise well to new data. To address this limitation, DTs can be combined.

Random Forest (RF) consists of multiple DTs and combines their predictions to improve performance and reduce overfitting [41]. Each tree in the RF is trained on a random subset of the data, selected with bootstrapping. At each node, a random subset of features is considered for splitting, further introducing diversity among the trees. The final prediction is made by aggregating the predictions of all the trees, typically through majority voting in the case of classification tasks. RFs can handle large volumes of data, leading to more accurate class predictions; however, this advantage often comes at the cost of increased computational time and resource consumption.

Naive Bayes (NB) is a supervised machine learning algorithm based on Bayes' theorem [42]. Similar to all previously covered algorithms, it is also used for classification tasks. NB makes the "naive" assumption that all the features used for class prediction are conditionally independent. It also assumes that each input contributes equally to the outcome. For classification, NB assigns the class with the highest probability based on the input features. The probability of a class given the features is calculated using Bayes' Theorem:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Where $P(C|X)$ is the probability of a class C given the features X, $P(X|C)$ is the likelihood of observing the features X given the class C, P(C) is the prior probability of the class C and P(X) is the probability of the features. One of its advantages is that it is computationally efficient and works well with high-dimensional datasets. While its assumptions overall perform well, the assumptions do not always hold, leading to incorrect classifications.

2.6 MyoCaptis Device

The MyoCaptis is a device designed to record EMG signals. It supports up to eight channels and includes two tactile grids capable of vibration feedback. Signal acquisition is performed using single-use bipolar surface electrodes. The device features a rechargeable battery and connects wirelessly via Bluetooth. The signal acquisition is performed through MATLAB, and the device supports sampling frequencies of 250 Hz, 500 Hz, 1000 Hz, and 2000 Hz. The device is shown in Figure 2.2, where the white cable represents the ground electrode, and the black and blue cables represent a bipolar electrode pair.

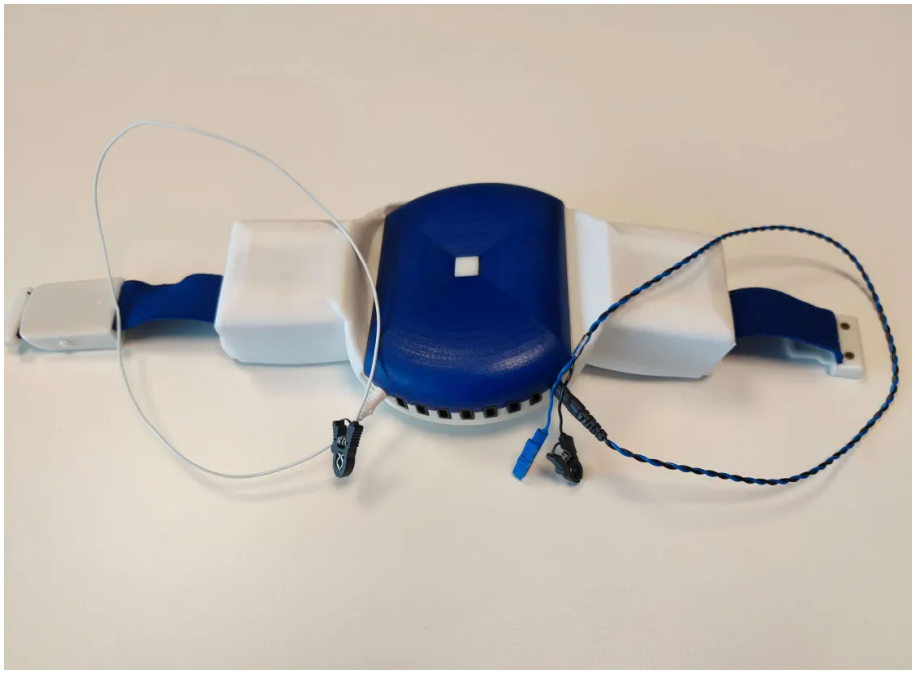


Figure 2.2: The Myocaptis device

3

Methods

This chapter describes the methodology developed and followed throughout the project. A literature review was initially conducted to identify common bio-signal contaminants, relevant signal features and classification models. EMG data were acquired from publicly available databases and systematically contaminated with noise. All signal features from the literature review were extracted from contaminated and non-contaminated signals. The features from contaminated and non-contaminated signals were statistically evaluated. Multiple classification models were trained and compared, leading to the selection of the most accurate model. The influence of segment length on classification performance was then assessed. Based on these findings, a single-label and a multi-label classification model were trained and tested. Lastly, a real-time implementation was created and validated to assess the feasibility of the proposed method in practice. All results can be found in Chapter 4.

3.1 Literature Review

The literature review explored available research on the identification and classification of EMG signal contaminants. The databases presented in the following list constituted the literature review.

- PubMed
- Scopus
- WebOfScience
- Springer

To enhance the search process and ensure the relevance of the results, keywords and MeSH terms were combined using Boolean operators. The specific search terms utilised are listed in Table 3.1. Terms from different categories were combined with the Boolean operator 'AND', while terms within the same category were combined using 'OR', allowing for variations within the defined categories.

Table 3.1: Literature review search terms by category

Category	Search Terms
Artefact	Anomaly, Artefacts, Noise
Classification	Classification, Identification, Noise Identification
Contamination	Contaminant, Contaminated, Contamination
EMG	Electromyogram, Electromyography, EMG, sEMG

Additionally, the following criteria were used to select relevant articles:

1. **Language**: The article must be in English.
2. **Type**: Only articles, conference papers, books, journals, or original theses will be considered.
3. **Availability**: Full-text access must be available either through public sources or through the Chalmers Library.

The bibliographies of the selected articles were thoroughly examined to identify additional relevant studies. In addition to including articles directly related to EMG, the literature review also included studies on contamination classification in ECG and EEG signals.

3.2 Data Acquisition

For the identification and classification of EMG contaminants, authentic non-contaminated EMG data were fundamental. For this purpose, several publicly available datasets were used from the databases in Table 3.2.

Table 3.2: Overview of databases used for EMG signal acquisition and their characteristics

Database	Subjects	Exercises	Channels	Muscles	Reference
MENDELEY	40	1	4	Forearm	[43]
META EMG2POSE	15	1	16	Wrist	[44]
NINAPro DB2	40	3	12	Forearm	[45]
NINAPro DB7	20	2	12	Arm	[46]

All databases had EMG signals with a sampling frequency of 2000 Hz, but with different pre-processing methods. Due to the large size of the full META EMG2POSE dataset (193 subjects), only a provided subset of 15 subjects was utilised. In addition to the individual datasets, a combined dataset was created by merging all available data.

3.2.1 Contamination

EMG signals were contaminated with three different contaminants: ECG, MA and WGN. All patients, exercises, and available channels were included in the contam-

ination process. Contamination was implemented in MATLAB by directly adding the contamination signals to the authentic EMG recordings, allowing for control of contamination type and intensity.

- **ECG**: For ECG contamination, six ECG recordings were utilised, each comprising four channels and a duration of eight seconds. The recordings were obtained from the MAECGDB database available on PhysioNet [47, 48]. For each contamination instance, a random ECG recording and channel were selected. To align with the EMG data, the ECG recordings were upsampled from 500 Hz to 2000 Hz. The upsampling was done using MATLAB's `resample()` function, which increases the sampling rate by interpolating new data points and applying an anti-aliasing filter.

- **MA**: For MA contamination, an MA recording with two channels was used. The MA recording was available at PhysioNet's database NSTDB [49, 48]. Due to its long duration, random segments were extracted and utilised for contamination. The original sampling rate of 360 Hz was upsampled to 2000 Hz, done using MATLAB's `resample()` function.

- **WGN**: For WGN contamination, WGN were created using MATLAB's `randn()` function, generating samples from a standard normal distribution, providing an approximation of a white spectrum.

For every clean EMG signal, seven contaminated versions were generated:

- EMG signal contaminated with ECG
- EMG signal contaminated with MA
- EMG signal contaminated with WGN
- EMG signal contaminated with ECG + MA
- EMG signal contaminated with ECG + WGN
- EMG signal contaminated with MA + WGN
- EMG signal contaminated with ECG + MA + WGN

Each combination was contaminated at five different SNR levels, ranging from -20 to 20 dB in 10 dB increments. For each SNR level, a copy of the clean signal was created to maintain perfect class balance. The noise types were added together and then scaled to the desired SNR. The scaling procedure began with the calculation of the mean power of the EMG signal segment. The power of the total noise signal was subsequently computed to determine the necessary scaling factor required to achieve the target SNR. Following this, the noise signal was scaled appropriately and added to the EMG signal. An example of EMG signals contaminated with WGN at different SNR levels is shown in Figure 3.1.

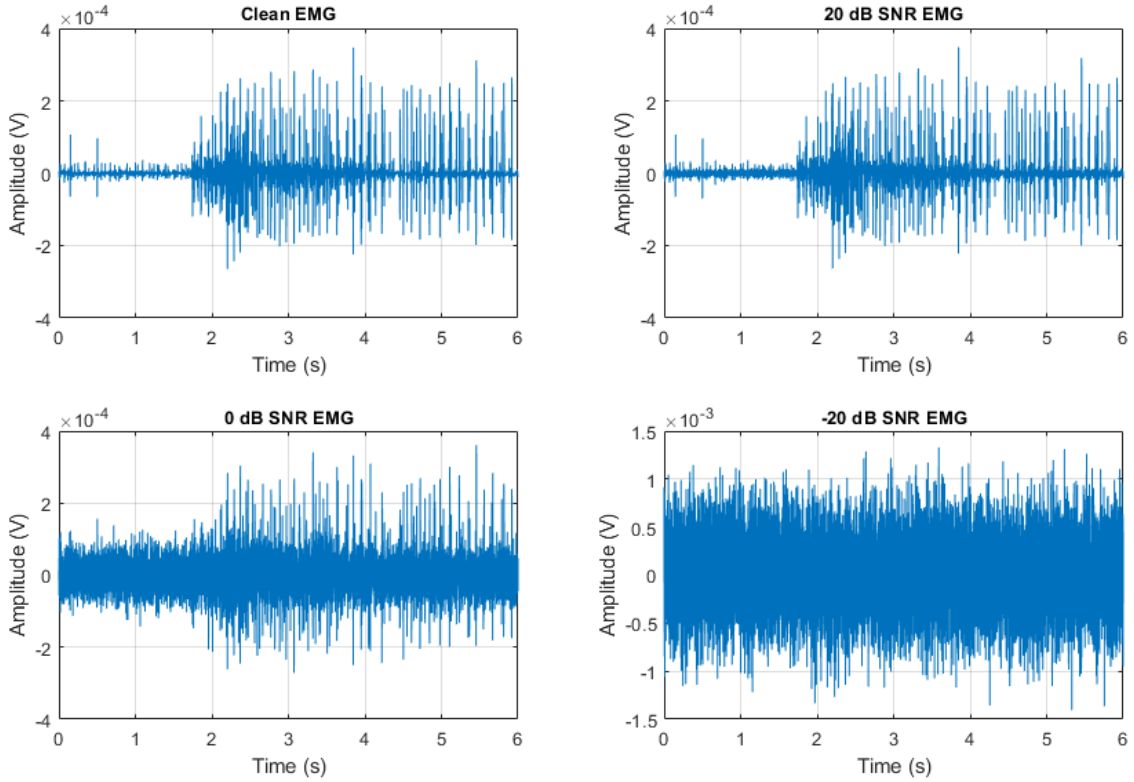


Figure 3.1: Non-contaminated EMG signal and EMG signal contaminated with WGN artefact at 20 dB, 0 dB, and -20 dB SNR

3.3 Feature Extraction

For the evaluation of signal features, and later the creation of training and test sets, signal features were extracted from both contaminated and non-contaminated EMG signals; all combinations of contaminants and SNR levels included.

Time domain features were calculated directly from the EMG signal. Frequency domain features utilised a sliding window approach, where each signal was segmented into overlapping windows of 1024 samples with 50% overlap. A Fast Fourier Transform were applied on each window, and the PSD was estimated using Welch's method. The method estimated the signal's frequency content by splitting the signal into overlapping segments, analysing each part, and averaging the results. This smoothed out random spikes or short-term noise, highlighting the overall frequency content. The resulting PSD data served as the basis for computing all frequency domain features.

All features were independently extracted for each EMG channel. The results were stored in structured tables, each row representing an EMG channel from a subject and exercise, and each column corresponding to a specific signal feature.

3.4 Signal Feature Evaluation

The process of selecting signal features for classification began with gathering all signal features identified in the literature review. Once identified, an initial screening was performed to exclude features that did not meet the criteria listed below.

- Not overlapping with other features
- Don't require knowledge of contaminants

Following the initial screening, the remaining features were extracted from contaminated and non-contaminated EMG signals. The signals were acquired as described in Chapter 3.2, and the signal features were extracted as described in Chapter 3.3. The extracted signal features formed a dataset for further evaluation.

One-way ANOVA tests were performed to select signal features for the feature set. Only features that showed statistically significant differences between signals with different types of contamination were included. The analysis began with signals from the NINAPro DB2 dataset at an SNR level of -20 dB. For features with p-values below 0.001, the SNR was incrementally increased by 10 dB from -20 to 20 dB. At 20 dB, features with p-values below 0.001 were selected for further evaluation.

After excluding signal features with a p-value above 0.001 for 20 dB SNR signals from the NINAPro DB2 dataset, One-way ANOVA tests were conducted on features from 20 dB signals in the remaining three datasets (NINAPro DB7, MENDELEY, META EMG2POSE). Finally, a One-way ANOVA test was performed on features calculated from signals from the combined dataset, accounting for variation between signals.

The final selection of signal features for the feature set included those with a p-value of at most 0.001 for at least three of the four individual datasets, and for the combined dataset. Out of the 41 features initially considered, 7 features were selected.

The signal features evaluated as the most promising from the feature evaluation in Chapter 3.4 were extracted from all data files prepared in 3.2. These extracted features form the foundation for the evaluation and training of classification models.

3.5 Model Evaluation

The process of selecting the most suitable model began with gathering model types used for contaminant classification in EMG signals from the literature review. Models compatible with signal features were noted for further evaluation.

The initial evaluation of the model utilised data from the NINAPro DB2 database. A flowchart outlining the process for the initial model evaluation is presented in Figure 3.2. The signals were acquired and contaminated as described in Chapter 3.2, but only non-contaminated and single-contaminated signals were included. The

signal features were extracted from seven minutes of data per subject and exercise, as described in Chapter 3.3.

The full range SNR utilised non-contaminated signals and signals with SNR levels between -20 dB and 20 dB, while the 20 dB datasets only utilised the non-contaminated signals and the signals with an SNR level of 20 dB. The two datasets allowed for evaluation of overall performance as well as more specific evaluation of performance on the highest SNR level, the most challenging classification task as described in the literature review.

Each model type was evaluated using five different train/test splits, each with a 70/30 ratio. For each split, the classification accuracy was calculated. The model types were evaluated based on the mean classification accuracy and standard deviation.

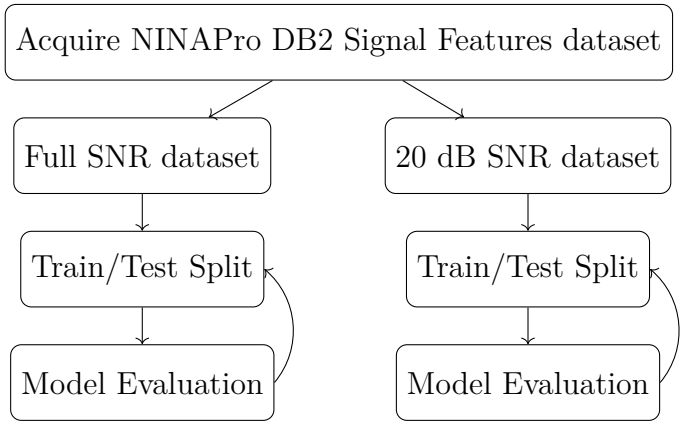


Figure 3.2: Flowchart for initial evaluation of machine learning models for classification of EMG contaminants for SNR levels: -20 dB to 20 dB, and 20 dB

Following the evaluation on NINAPro DB2, the three model types with the highest total accuracy were further evaluated. The model types were evaluated on the remaining datasets: NINAPro DB7, MENDLEY, META EMG2POSE, and the combined dataset. All four datasets underwent the same approach as NINAPro DB2, as presented in Figure 3.2, using the same five-iteration approach. The model with the highest overall mean accuracy across all datasets was selected.

For further insight, the final model underwent an additional evaluation step. It was trained on the combined dataset with the full SNR range and tested on individual SNR levels (-20 dB, -10 dB, 0 dB, 10 dB, 20 dB) to assess classification accuracy for different contamination levels. For each SNR level, the five-iteration approach was utilised for evaluation.

3.6 Segment Length Evaluation

The influence of window length on classification accuracy was initiated by creating datasets with segments of each desired length: 1 second (s), 2 s, 3 s, 4 s, 5 s, 6 s, 7 s,

8 s, 9 s, and 10 s. Each dataset included five segments per subject and exercise from all databases; each segment was contaminated according to the procedure described in Chapter 3.2.1 and the signal features extracted as described in Chapter 3.3.

Once separate datasets with window lengths of 1 to 10 s were established, the impact of window length on classification accuracy was studied. The evaluation was performed over five iterations, dividing the datasets into 70/30 train/test splits, training and testing a single-label RF model for each iteration. The classification accuracy was presented as mean accuracy and standard deviation. The models were trained on a full range SNR version of the dataset, and tested on the full SNR range and 20 dB, respectively, for each iteration. The choice of window lengths was made by comparing the mean classification accuracy and standard deviation between different window lengths.

To assess the impact of segment length, the F-statistic was employed to evaluate the discriminatory power of the signal features for each window length. The F-statistic measures the ratio of the variance between classes (Clean, ECG, MA, WGN) to the variance within each class, with higher values indicating features that more effectively differentiate between the categories.

3.7 Single- and Multi-label Classification

After determining the segmentation length of five seconds, as described in Chapter 3.6, the final dataset was assembled. All databases, subjects, exercises, and channels from Table 3.2 were utilised. The first 12 segments from all available datasets were contaminated according to Chapter 3.2.1, and the features were extracted as described in 3.3. All combinations of contaminants and SNR levels were included, labelling the files by present contamination or non-contamination.

Two versions of the best-performing model architecture were trained on the dataset: one single-label and one multi-label classification model, as described below. Both models were trained and tested using a 70/30 split over five iterations, training and testing a new model on a new split for every iteration. The models were evaluated by accuracy and standard deviation.

- **Single-label model**, categorising data into one of four categories: Clean, ECG, MA, or WGN. The model is trained and tested on the full SNR range, on signals with one type of contamination or non-contaminated signals, without combinations of contaminants.
- **Multi-label model**, categorising data into one or more categories, allowing for combinations. The model is trained and tested on the full SNR range, signals with one type of contamination, non-contaminated signals, and with combined contaminations.

3.8 Online Implementation

A real-time implementation was developed using the MyoCaptis device to assess single- and multi-label model feasibility in a practical setting.

Figure 3.3 illustrates the real-time classification workflow. The models were exported to MATLAB, where a script was developed to process live data in batches of five seconds from MyoCaptis, pre-process the signal, extract signal features, and perform real-time classification. The pre-processing consists of a 50 Hz Notch filter followed by a 15-500 Hz Bandpass filter. After each prediction, the script updates a live plot to visualise the EMG signal. This allows for visual inspection to ensure that the signal appears reasonable.

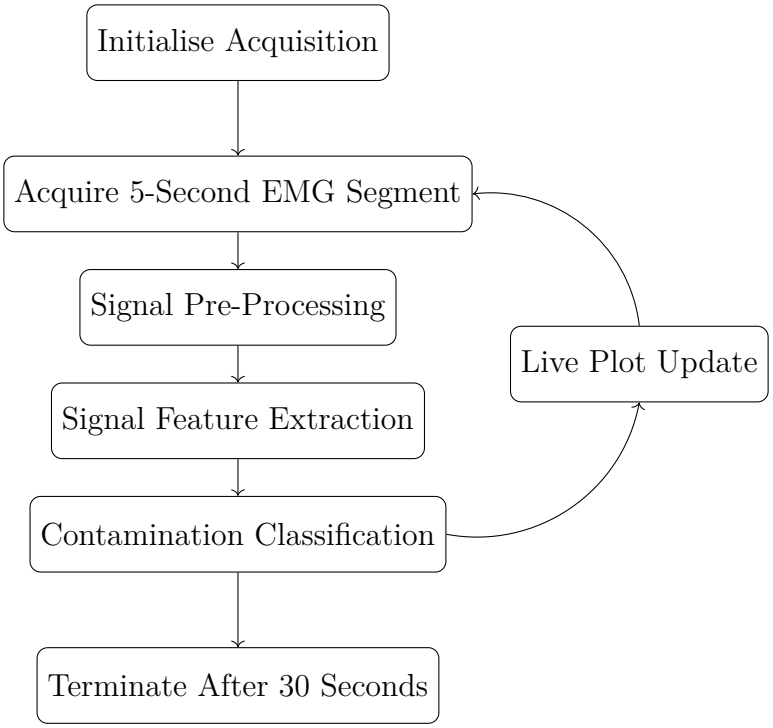


Figure 3.3: Flowchart of the real-time classification pipeline implemented with the Myocaptis device

3.8.1 Online Validation

The single- and multi-label models were validated online via the real-time implementation. The EMG signals were recorded live from a MyoCaptis user's forearm, the Brachioradialis muscle, and a reference on the elbow. Pre-recorded contaminants were scaled for SNR-specific evaluation and added to the EMG signal.

Both models were evaluated during six five-second segments per SNR level, activity, and contaminant. EMG signals with SNR levels -20 dB, -10 dB, 0 dB, 10 dB, and 20 dB were evaluated during the activities: resting, and periodical activity.

Each contaminant was recorded separately. For ECG interference, an electrode was placed below the collarbone on the user's hand side to record an ECG signal. MA was recorded by attaching the electrodes to the forearm and manually disturbing the skin–electrode interface by tugging the electrodes and moving the skin around them. WGN was recorded by creating a loop with electrode wires and moving a strong magnet through the loop to induce broadband electrical noise. The contamination recordings were segmented into five-second windows.

During the evaluation were live EMG recordings in batches of five seconds were contaminated with scaled contamination recordings. Each contaminated segment was classified by the single-label and the multi-label model. After six segments, the accuracy was calculated, and a majority vote was performed. The process was repeated for each contaminant type, SNR level, and activity. During resting segments, the forearm rested flat on a table, and during the active segments, the forearm rested on the table, and the wrist was flexed upwards for two seconds and then rested for two seconds.

Performance was assessed in two ways:

1. **Per-segment accuracy**, in which each individual five-second segment is classified separately. Accuracy is then calculated as the proportion of correctly classified segments relative to the total number of segments.
2. **Majority voting**, in which each individual five-second segment is classified separately, and the most frequently predicted label is chosen as the classification label. Accuracy is either 0 or 1.0.

4

Results

This chapter presents the results of this project, structured into three key areas: Literature Review, Feature Evaluation, Model Evaluation, Segment Length Evaluation, Single- and Multi-Label Classification, and Real-Time Implementation.

4.1 Literature Review

The number of search results obtained using the search terms described in Chapter 3.1 is presented in Table 4.1 below, obtained before 7/2-2025. One column presents the total number of articles found, the other column after the three additional criteria listed below Table 3.1 in Chapter 3.1. Many articles appear in multiple databases, resulting in overlap between sources. The bibliographies of all articles were screened for additional articles relevant to the study questions, and relevant articles were included in the literature review.

Table 4.1: Literature review search result counts before and after screening of abstracts

Database	Count Before Screening	Count After Screening
PubMed	29	8
Scopus	63	11
Springer	50	0
WebOfScience	53	6

4.1.1 EMG Contamination

Table 4.2 presents all EMG signal contaminants identified in at least two articles from the literature review. The three most common EMG contaminations are ECG interference, Motion Artefact, and White Gaussian Noise. Power Line Interference is excluded due to its narrow power spectrum, its removal not affecting the EMG information as much as the removal of other contaminants.

Table 4.2: EMG signal contaminants identified in a minimum of 2 articles from the literature review

Contaminant	References
Amplifier Saturation	[6], [11], [14], [15], [28], [27]
Analog-to-Digital Converter Clipping	[6], [11], [27], [28]
Baseline Wander	[6], [10], [18], [20], [50]
Cross-talk	[6], [28], [50]
ECG Interference	[6], [7], [8], [9], [11], [12], [13], [14], [15], [16], [17], [19], [20], [27], [28], [50]
Motion Artefact	[6], [7], [8], [11], [13], [14], [15], [16], [17], [18], [19], [20], [27], [28], [50]
Power Line Interference	[6], [7], [8], [9], [10], [11], [13], [14], [15], [16], [17], [19], [20], [27], [28], [50]
White Gaussian Noise	[7], [8], [10], [13], [14], [16], [17], [50]

The contaminations of EMG signals investigated in the literature are naturally occurring or artificially introduced. The origins of these contaminations, as identified in the literature review, are presented in Table 4.3.

Table 4.3: Source of EMG contamination identified in the literature review

Contamination Source	References
Intrinsic Contaminants	[18], [28]
Recorded Contaminants	[7], [9], [12]
Simulated Contaminants	[7], [8], [11], [13], [14], [15], [16], [17], [18], [19], [20], [27], [28]

4.2 Signal Feature Evaluation

The signal features presented in Table 4.4 achieved a p-value below 0.001 in the ANOVA test conducted across the classes: Clean, ECG, MA, and WGN. Table 4.4 presents the p-values from ANOVA tests conducted on each individual dataset and

the combined dataset, with the selected signal features highlighted in italic. The tests were performed on signals with an SNR of 20 dB.

Table 4.4: P-values for signal features for all individual datasets for SNR of 20 dB. Features highlighted in italic were selected based on achieving a p-value below 0.001 in at least three of the four individual datasets and the combined dataset

	META EMG2POSE	**MEND-ELEY**	**NINAPro DB2**	**NINAPro DB7**	**Combined**
Hurst Exponent	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
L-Kurtosis	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Median Frequency	0.233	0.792	< 0.001	< 0.001	< 0.001
Mean Frequency	0.003	0.115	< 0.001	< 0.001	< 0.001
Negentropy 1	0.009	0.011	< 0.001	0.009	< 0.001
Pearson CCN	0.355	0.764	< 0.001	< 0.001	< 0.001
PSD Deformation	0.016	< 0.001	< 0.001	< 0.001	< 0.001
PSD Deformation Ratio	< 0.001	0.318	< 0.001	< 0.001	< 0.001
Power Spectrum Range	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Shannon Entropy	0.367	0.011	< 0.001	0.009	0.039
Slope Sign Changes	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Spectral Entropy	0.138	0.927	< 0.001	< 0.001	< 0.001
Zero Crossings	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

4.3 Model Evaluation

All models used in the literature for classifying signal contaminants are presented in Table 1.2. Among these, the models listed below are suitable for signal features.

- Decision Tree
- K-Nearest Neighbour
- Linear Discriminant Analysis
- Naive Bayes
- Quadratic Classifier
- Random Forest
- Support Vector Machine

Table 4.5 presents the classification accuracy of the seven models trained and tested on the NINAPro DB2 dataset. The accuracy is reported for the full SNR range (-20 dB to 20 dB) and at only 20 dB. Results are expressed as the mean accuracy with standard deviation across five classification rounds. The three models (DT, LDA, and RF) with the highest total accuracy, across the full SNR range and 20 dB SNR, were selected for further investigation.

Table 4.5: EMG contamination classification accuracy for different model architectures for two different SNR ranges: -20 db to 20 dB, and 20 dB

Model	Training SNR Range	Testing SNR Range	Accuracy
DT	-20 dB → 20 dB	-20 dB → 20 dB	0.915 ± 0.005
KNN	-20 dB → 20 dB	-20 dB → 20 dB	0.757 ± 0.004
LDA	-20 dB → 20 dB	-20 dB → 20 dB	0.859 ± 0.004
NB	-20 dB → 20 dB	-20 dB → 20 dB	0.708 ± 0.001
QC	-20 dB → 20 dB	-20 dB → 20 dB	0.888 ± 0.004
RF	-20 dB → 20 dB	-20 dB → 20 dB	0.959 ± 0.001
SVM	-20 dB → 20 dB	-20 dB → 20 dB	0.579 ± 0.002
DT	20 dB	20 dB	0.779 ± 0.004
KNN	20 dB	20 dB	0.590 ± 0.013
LDA	20 dB	20 dB	0.736 ± 0.006
NB	20 dB	20 dB	0.563 ± 0.013
QC	20 dB	20 dB	0.697 ± 0.026
RF	20 dB	20 dB	0.799 ± 0.009
SVM	20 dB	20 dB	0.339 ± 0.003

Tables 4.6 and 4.7 present the classification accuracies of the three best-performing models from Table 4.5, for all datasets. Classification accuracy and standard deviation were calculated over five iterations for each individual dataset and for the combined dataset. Table 4.6 reports the results for the full SNR range (-20 dB to 20 dB), while Table 4.7 focuses on an SNR of 20 dB. RF achieved the best overall performance.

Table 4.6: EMG contamination classification accuracy for SNR range -20 dB to 20 dB per dataset for the top three models

	MENDELEY	META EMG2POSE	NINAPro DB2	NINAPro DB7	Combined
DT	0.861 ± 0.012	0.967 ± 0.004	0.915 ± 0.005	0.905 ± 0.004	0.892 ± 0.005
LDA	0.821 ± 0.010	0.766 ± 0.003	0.859 ± 0.004	0.846 ± 0.004	0.744 ± 0.003
RF	**0.904 ± 0.008**	**0.979 ± 0.002**	**0.959 ± 0.001**	**0.945 ± 0.005**	**0.958 ± 0.002**

Table 4.7: EMG contamination classification accuracy for 20 dB SNR per dataset for the top three models

	MENDELEY	META EMG2POSE	NINAPro DB2	NINAPro DB7	Combined
DT	0.551 ± 0.042	0.939 ± 0.003	0.779 ± 0.004	**0.738 ± 0.009**	0.784 ± 0.002
LDA	**0.601 ± 0.023**	0.790 ± 0.009	0.736 ± 0.006	0.709 ± 0.021	0.647 ± 0.005
RF	0.503 ± 0.036	**0.962 ± 0.010**	**0.799 ± 0.009**	0.711 ± 0.013	**0.799 ± 0.005**

Table 4.8 presents the mean classification accuracy with standard deviation from five rounds of classification per SNR for the best-performing model from Tables 4.6 and 4.7.

Table 4.8: EMG contamination classification accuracy for best performing model per SNR level, trained on the combined dataset with SNR range -20 dB to 20 dB

SNR [dB]	Accuracy
-20	1.000 ± 0.004
-10	0.999 ± 0.0004
0	0.995 ± 0.001
10	0.964 ± 0.003
20	0.836 ± 0.003

4.4 Segment Length Evaluation

Figure 4.1 illustrates the changes in classification accuracy for the single-label RF model trained on the full dataset and tested under two different conditions: the full SNR range and an SNR of 20 dB, across window lengths ranging from 1 to 10 seconds. The mean accuracy and standard deviation are calculated over five classification iterations.

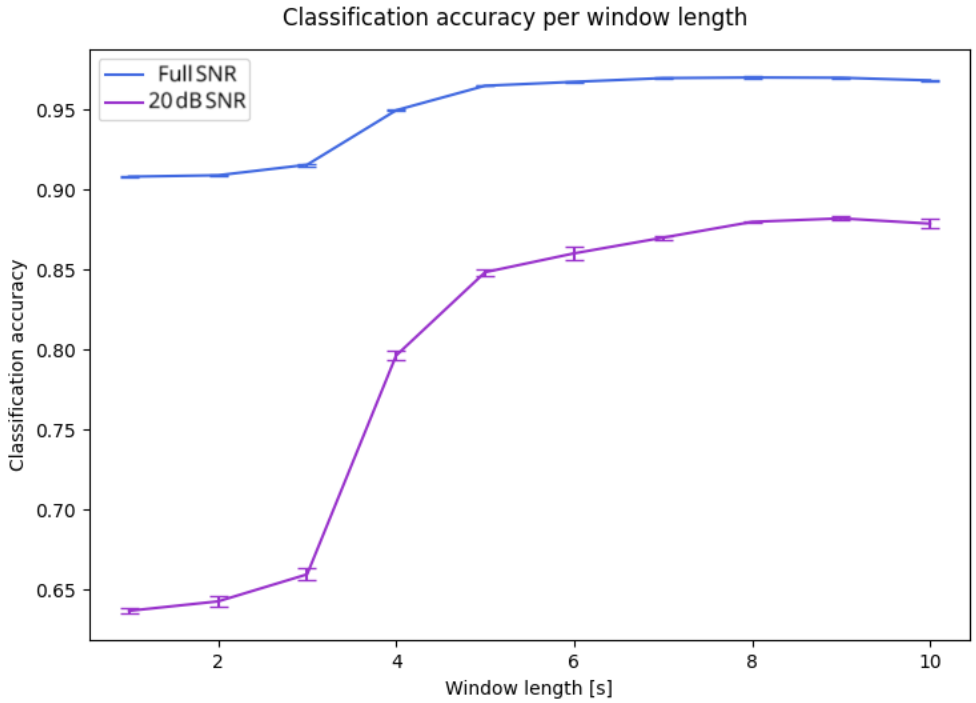


Figure 4.1: Effect of segment window length (1–10 seconds) on classification accuracy at full SNR range and at 20 dB SNR

Figure 4.2 visualises the F-statistic values for each signal feature across different window lengths, per dataset. Figure 4.3 presents the normalised F-statistic values; the normalisation is performed by the mean value per feature and dataset.

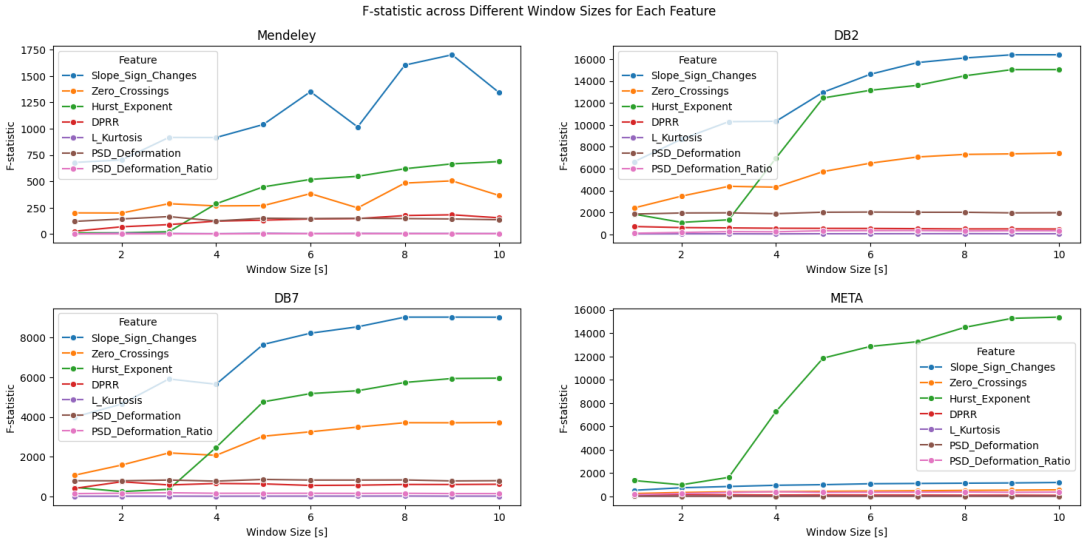


Figure 4.2: F-statistic value for each signal feature at different window lengths, across datasets from MENDELEY, NINAPro DB2, NINAPro DB7, and META EMG2POSE

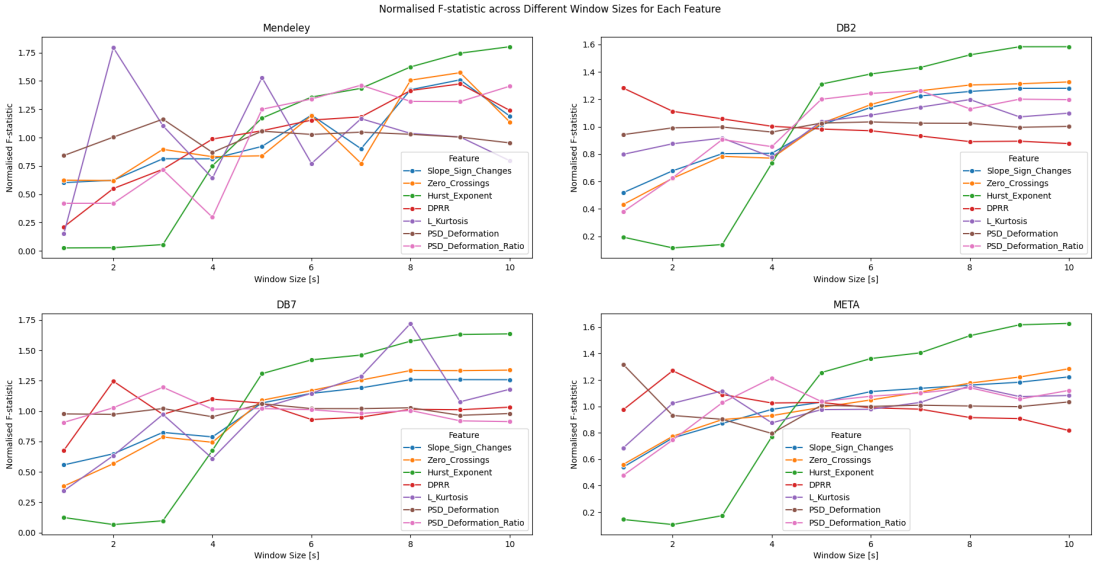


Figure 4.3: Normalised F-statistic value for each signal feature at different window lengths, across datasets from MENDELEY, NINAPro DB2, NINAProDB7, and META EMG2POSE

4.5 Single- and Multi-label Classification

Two models were trained and evaluated: a single-label RF model, and a multi-label RF model. Table 4.9 presents the classification accuracy for segmentation lengths of five seconds, calculated over five iterations, as mean accuracy and standard deviation.

Table 4.9: EMG signal contamination classification accuracy over five iterations for single-label model and multi-label model for full SNR range

Model	Classification Accuracy
Single-label Model	0.967 ± 0.0003
Multi-label Model	0.767 ± 0.0005

The confusion matrix for the single-label model from five iterations of the full SNR range is visualised in Figure 4.4. The confusion matrix for the multi-label model from five iterations for the full SNR range is visualised in Figure 4.5.

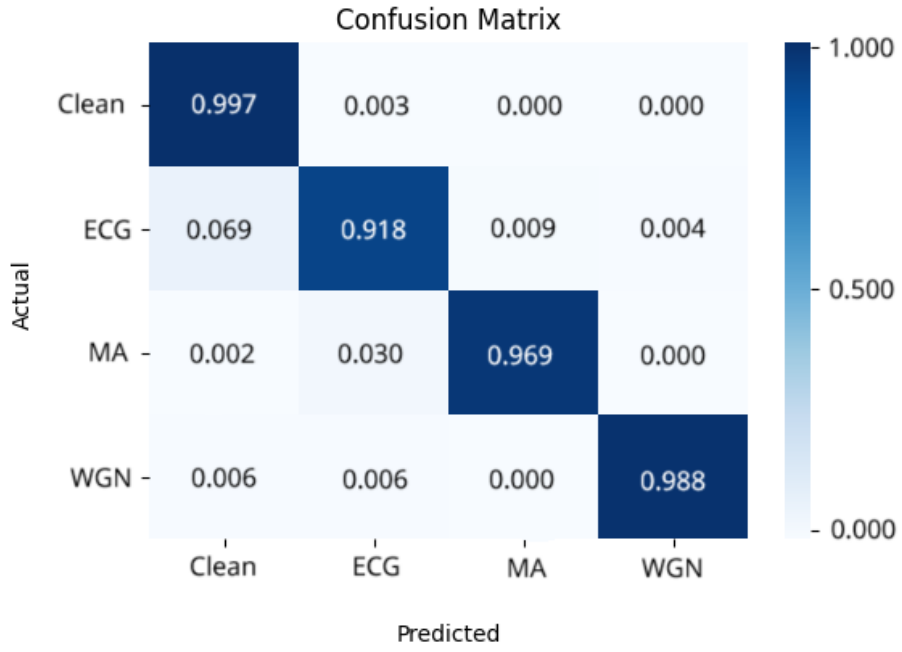


Figure 4.4: Confusion matrix for EMG signal contamination classification by single-label model over five iterations

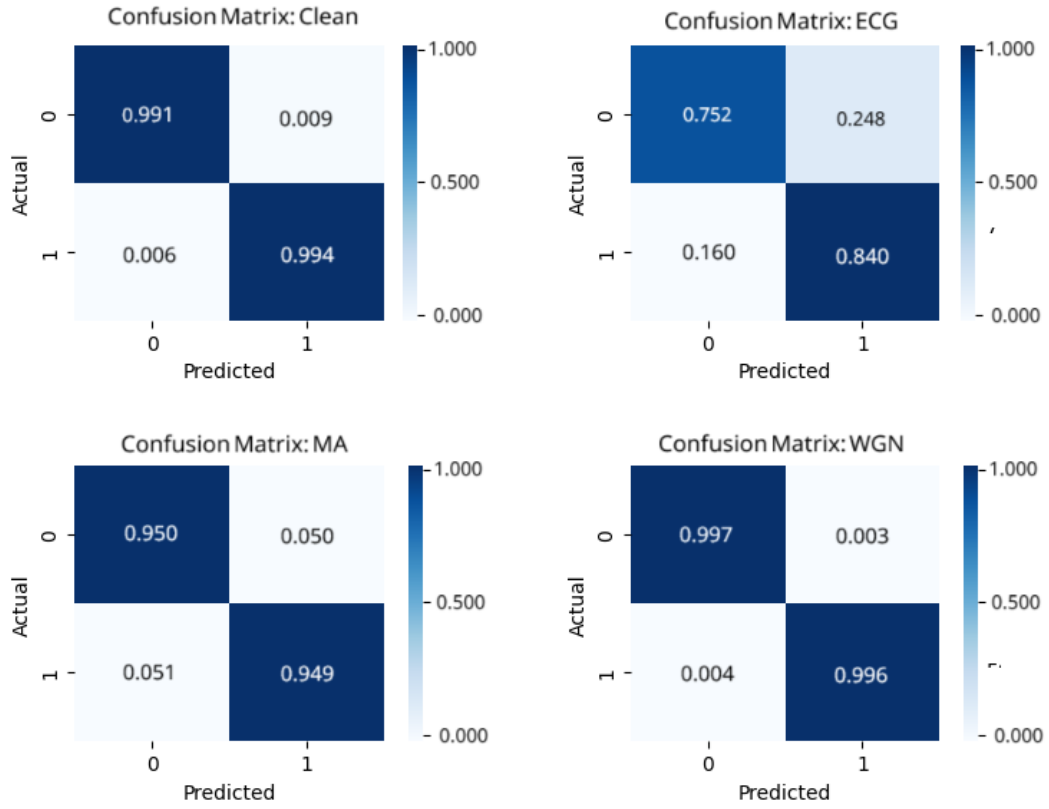


Figure 4.5: Confusion matrix for EMG signal contamination classification by multi-label model over five iterations

4.6 Online Validation

Tables 4.10, 4.11, and 4.12 present the results for online validation performed for the three contaminations: ECG, MA, and WGN. For each SNR level, activity, and contaminant, six five-second segments were classified. The 'Accuracy' columns show the number of correct classifications out of all six segments. The Majority Voting takes the classification of all segments into account. In case of a tie, both of the labels are presented. For the multi-label model, only whether the target contaminant was correctly classified is considered when calculating accuracy. For instance, if only ECG is present in the signal and the model predicts both ECG and MA, the accuracy is calculated as 100%, but the majority vote will say "ECG/MA".

Table 4.10: Online validation results of single- and multi-label models for classification of ECG contaminants in EMG signals across SNR levels -20 dB to 20 dB. Results presented for both classification accuracy and major voting

		Single-label Model		**Multi-label Model**	
SNR [dB]	**Activity**	Accuracy	Majority Vote	Accuracy	Majority Vote
-20	Rest	1.000	ECG	1.000	ECG
-10	Rest	1.000	ECG	1.000	ECG
0	Rest	1.000	ECG	1.000	ECG
10	Rest	0.833	ECG	0.833	ECG
20	Rest	0.000	Clean	0.000	Clean
-20	Active	1.000	ECG	1.000	ECG
-10	Active	1.000	ECG	1.000	ECG
0	Active	1.000	ECG	1.000	ECG
10	Active	0.833	ECG	1.000	ECG
20	Active	0.833	ECG	0.666	ECG

Table 4.11: Online validation results of single- and multi-label models for classification of MA contaminants in EMG signals across SNR levels -20 dB to 20 dB. Results presented for both classification accuracy and major voting

		Single-label Model		**Multi-label Model**	
SNR [dB]	**Activity**	Accuracy	Majority Vote	Accuracy	Majority Vote
-20	Rest	1.000	MA	1.000	MA
-10	Rest	1.000	MA	1.000	MA
0	Rest	1.000	MA	1.000	MA
10	Rest	0.166	ECG	0.333	ECG
20	Rest	0.000	Clean	0.000	Clean
-20	Active	1.000	MA	1.000	MA
-10	Active	1.000	MA	1.000	ECG/MA
0	Active	1.000	MA	1.000	ECG/MA
10	Active	0.500	ECG/MA	1.000	ECG/MA
20	Active	0.000	Clean/ECG	0.000	Clean

Table 4.12: Online validation results of single- and multi-label models for classification of WGN contaminants in EMG signals across SNR levels -20 dB to 20 dB. Results presented for both classification accuracy and major voting

		Single-label Model		**Multi-label Model**	
SNR [dB]	**Activity**	Accuracy	Majority Vote	Accuracy	Majority Vote
-20	Rest	1.000	WGN	0.333	WGN
-10	Rest	0.000	Clean	0.000	Clean
0	Rest	0.000	Clean	0.000	Clean
10	Rest	0.000	Clean	0.000	Clean
20	Rest	0.000	Clean	0.000	Clean
-20	Active	1.000	WGN	1.000	WGN
-10	Active	1.000	WGN	1.000	WGN
0	Active	0.666	WGN	0.666	WGN
10	Active	0.000	Clean	0.000	Clean
20	Active	0.000	Clean	0.000	Clean

5
Discussions

This chapter discusses several key areas of this project, starting with the literature review, including contaminations and the role of SNR in classifier performance. Following the literature review, the evaluation of signal features is discussed, making way for the discussion of model evaluation, where aspects influencing model behaviour and limitations are examined. The impact of segment length on the classification performance is debated, followed by the performance differences between single- and multi-label models. The implications of the online validation results are also discussed, particularly model sensitivity to low-frequency noise. The chapter concludes with recommendations for future work.

5.1 Literature Review

Signal contaminations are widely studied in EMG research, but not all types of noise are equally relevant for practical, real-time applications. Table 4.2 presents the frequency of EMG contaminants discovered during the literature review. The five most common contaminants are: Amplifier Saturation, ECG Interference, Motion Artefact, Power Line Interference, and White Gaussian Noise. More technical artefacts, such as Amplifier Saturation and Analogue-to-Digital Converter Clipping, are common but difficult to mitigate during use and better addressed through hardware modifications. Power Line Interference is frequently addressed in the literature. It is well-known and removable, and was therefore not prioritised. With the aim to choose contaminants most commonly encountered in clinical and physiotherapy settings, as well as possible to remedy during use, the chosen contaminants were ECG interference, MA, and WGN.

The three approaches for the generation of contaminations are presented in Table 4.3; utilising simulated contaminants is the dominant approach in the reviewed literature. With a lot of information available on the simulation of contaminants, artificial contamination was employed. This choice enabled controlled manipulation of SNR levels. However, it is important to note that this method represents an idealised scenario. Unlike real-world contaminations, artificial noise can lack the variability and complexity of naturally occurring artefacts. This approach supports consistent experimentation, but it can limit the generalisability of results to real-world applications.

SNR plays a critical role in classification performance. In all articles presenting SNR

values, the classification accuracy rapidly decreases with increasing SNR, as exemplified in Table 1.4. Achieving reliable performance at around 20 dB SNR proved particularly difficult. The challenge is logical, since the ratio between contamination and EMG signal increases with a decreasing SNR, the features differentiating between contaminants become more prominent with decreasing SNR. Signals with high SNR are more ambiguous and therefore pose a greater challenge. The example from Table 1.4 presents a steep decline in accuracy as the SNR goes from -10 dB, to 0 dB. Figure 3.1 illustrates the effect of SNR level for -20 dB, 0 dB, and 20 dB. At 0 dB, the amount of EMG signal and contamination are equal, and the signal quality is significantly affected by the contaminant. Generally, the higher SNR level, the higher the quality of the EMG signal. With the knowledge of present contaminations, filters can be adapted to more efficiently remove present contamination, increasing signal quality and SNR. To provide a model robust enough for consistent use in clinical settings, better classification accuracy must be presented for a wider range of SNR levels to allow for further improvement of EMG signal quality.

5.2 Signal Feature Evaluation

Table 1.1 from Chapter 4.1 summarises the different strategies for extracting data from EMG signals utilised in the literature. Most articles use signal features, and a lot of information is available regarding different features. Signal features are also versatile; an extensive range of information can be extracted from both the time and frequency domains. Compared to Signal-ICA components and time-series signals, the extracted signal features require significantly less memory and computational resources, possibly making them more suitable for faster implementations.

The total number of signal features presented in the literature review sums up to 41 and is presented in Table A.1 and Table A.2 in Appendix A. Even though many articles have explored multiple signal features, and some have performed statistical comparisons, none have performed statistical evaluations on all features available in the literature. Some have performed visual evaluations rather than computer-based evaluations. Computer-based statistical evaluations are perhaps more appropriate, as they can identify subtle differences hidden from the human eye, especially given that the classification is conducted by a computer. The choice, therefore, fell on performing a statistical evaluation on all available signal features in the creation of the feature set.

The statistical method initially chosen for evaluation was p-values from an ANOVA test performed in Python. Table 4.4 presents the signal features that passed the initial feature evaluation. The signal features are presented with p-values for each dataset, allowing for an overview of how well that feature differentiates between different contaminations on different databases. Since the contamination applied is identical between all databases, the differences in p-value are due to variation in EMG between the databases. Some p-values vary greatly between the different databases. A trend is that the MENDELEY and META EMG2POSE are the worst-performing databases overall. It is known that ANOVA tests allow for more

outliers, the bigger the dataset. Due to the sheer size of the datasets, especially the combined dataset, it is up for discussion whether the low scores are representative or due to the data size. Perhaps another statistical measure should have replaced or complemented the p-value evaluation. The F-statistic measure was included as a tool to evaluate the effect of segment length on the discriminatory power of the dataset. Perhaps, F-statistic values should be included earlier and present an additional dimension of information for the evaluation of signal features.

5.3 Model Evaluation

The literature has evaluated many machine learning model architectures for EMG contaminant classification, as presented in Table 1.2. Most models perform well during low SNR levels but poorly when the SNR exceeds 0 dB. Since most models have been extensively studied, greater focus was placed on the choice of signal features. Performance may depend more on the nature of the data and the ability to distinguish between contaminants, rather than on the model architecture. Models with high computational load were excluded from the evaluation, hoping for a less computationally heavy model with the possibility of implementing a real-time solution.

Following the extensive evaluation of signal features, models compatible with signal features were evaluated. They were evaluated on two different SNR-datasets as presented in Table 4.5, and the data came from one database only (NINAPro DB2). One dataset had a full SNR range, from -20 dB to 20 dB, and another only the 20 dB SNR. This is to evaluate overall performance and performance close to a clean signal. As presented in Table 1.3, the literature review denoted signals with SNR above 15/18/20 dB as clean. Classification performance recorded in the literature declines as SNR increases. Therefore, the specific 20 dB dataset was utilised for model evaluation.

The three models with best performance from Table 4.5 were evaluated on data from all four databases, and the combined dataset, with two different SNR ranges: full as presented in Table 4.6 and 20 dB only as presented in Table 4.7. This allowed for the comparison of accuracy between the databases. At full-range SNR, all models showed relatively consistent performance across the datasets, with the RF model achieving the highest classification accuracy. At 20 dB SNR, however, performance varied more within and between the datasets, highlighting the challenges associated with classification signals with high SNR levels. MENDELEY is the most challenging database; possibly due to its hardware, electrode placement, the signal pre-processing or exercises performed. To see the full impact of SNR level on classification accuracy, a model was trained on the full SNR range for the combined dataset and tested on the individual SNR levels as presented in Table 4.9. The impact of SNR is lower than presented in the literature review, but still follows the same trend.

A consistent trend is observed regarding which datasets achieve the highest and lowest classification accuracies. Table 4.6 and Table 4.7 present the classification

accuracy for the top 3 performing models on all datasets. For both the full SNR range and the 20 dB SNR range, NINAPro DB2 and META EMG2POSE are classified with better accuracy, while DB7 and MENDELEY are classified with inferior accuracy. The exact cause of this discrepancy remains unclear, but several possible explanations can be considered. One factor may be the balance between datasets. As shown in Table 3.2, the datasets vary in size, providing the model with differing sizes of training data. Generally, a greater amount of data provides better conditions for classification. However, NINAPro DB2 and NINAPro DB7 are the largest datasets, so this alone does not fully explain the variance in accuracy. Another potential explanation regards the pre-processing of EMG signals; different pre-processing methods may influence the ability to discriminate between various types of contamination. Other factors possibly contributing to the difference in classification accuracy between the datasets are: the exercises performed and how the amplitude changes, hardware variations, and electrodes and their placement. All aspects could contribute to the performance trend, but none is probably solely responsible for the variation.

Table 4.9 presents the classification accuracy per SNR level. At 20 dB, 84 % of the contaminants were correctly classified, while the literature provides a much lower accuracy. The main reason for this difference is possibly the statistical evaluation of signal features, selecting only those features that effectively differentiate between EMG contaminants. Additionally, the model was trained and evaluated across multiple datasets, exposing it to a broader distribution of contamination scenarios and reducing the risk of overfitting to signal specifics or recording setup. Most articles used a single dataset for model training. By incorporating multiple datasets, the robustness of the model was increased, achieving better handling of EMG variability.

5.4 Segment Length Evaluation

The choice of segment length is a trade-off between accuracy and time. A longer segment provides more data and allows for a more reliable classification, while a shorter segment can detect changes faster, but at the cost of lower accuracy. The main reason for studying the performance of different segment lengths was to determine how much EMG data is required for accurate classification, aiming to find the optimal balance where segments are as short as possible while still achieving high classification accuracy. An approach that increases the classification frequency while using longer data segments involves using previous data to perform classification at shorter intervals. But this would increase the computational load and is not guaranteed to capture changes as rapidly as a shorter segment length.

Figure 4.1 visualises how the classification accuracy changes with segmentation window length. A large increase in accuracy occurs between three and four seconds, and continues somewhat between four and five seconds. The F-statistic values presented in Figure 4.2 and Figure 4.3 follow the same pattern, hinting that the change is due to changes in characteristics such as discriminatory power as the segment length increases. The different signal features follow this trend to varying extent,

hinting that some features perhaps affect the classification more than others, e.g. Hurst Exponent shows a very prominent trend of big changes between three and five seconds, for all four datasets. A possible explanation for this change is related to the underlying activity pattern: all databases generally follow a repeated sequence of three seconds of activity and three seconds of rest. The four-second segment is the first where all segments are guaranteed to include both rest and activity, which could explain the big change in accuracy between the three- and four-second window segments.

5.5 Single- and Multi-label Classification

The classification accuracy for the single-label model and the multi-label model is presented in Table 4.9; The single-label model outperforms the multi-label model by approximately 20 percentage points. This performance gap is likely due to the increased complexity inherent in multi-label classification. The single-label model is only required to identify a single dominant contaminant. In contrast, the multi-label model must simultaneously evaluate the signal between four labels (Clean, ECG, MA, and WGN). The model can misclassify the type of contamination, but also fail to detect a contaminant altogether. SNR is superposition-based; when multiple contaminants are present in an EMG signal, their contributions combine to form the total SNR. As a result, each contaminant effectively has a lower SNR than the overall signal. This effectively weakens the presence of each contaminant in the signal, reducing their distinguishability. This superposition of contaminants results in less distinct patterns, which may explain the decreased model performance.

The confusion matrix for the single-label model is illustrated in Figure 4.4. Most labels are classified correctly, creating a strong diagonal in the confusion matrix. Even though all labels have high classification accuracies, ECG proves to be the most challenging. The most common incorrect prediction for ECG labels is Clean; possibly, the EMG values with the highest SNR are incorrectly classified as Clean. The Clean EMG signals have the best classification accuracy. Generally, very few are incorrectly classified as WGN or MA.

The confusion matrix for the multi-label model is illustrated in Figure 4.5. The multi-label model consistently struggles with ECG contamination, often over-predicting its presence even when it is absent. This behaviour could be explained by the signal features' inability to efficiently separate the unique temporal structure of ECG artefacts from other characteristics, assuming non-contaminated signals to be contaminated with ECG interference. The model is therefore overly sensitive and prone to false positives for ECG interference. The models best result is achieved for WGN contamination, with only 0.7% incorrectly classified.

Some characteristics are shared between both models, while others differentiate the models from each other. The single-label model performs best on Clean EMG data, while the multi-label model performs best on WGN-contaminated EMG data. In both models, the Clean label is occasionally predicted for signals with a higher

SNR. This may indicate a limitation in the models' ability to detect low-intensity contamination or a potential bias in how non-contaminated the EMG signals from databases were. The most challenging contaminant for classification is ECG for both models, the multi-label model has a bigger challenge with the classification than the single-label model. Both models exhibit similar sensitivities, but the single-label model simplifies the task by enforcing mutual exclusivity between classes, helping performance but risking ignoring secondary contaminants. In comparison, the multi-label model allows for richer representations of complex signals, but struggles more with overlapping noise patterns.

5.6 Online Validation

The results from the online evaluation of single- and multi-label models are presented in Table 4.10 for ECG contamination, Table 4.11 for MA contamination, and Table 4.12 for WGN contamination. Compared to the classification results for single- and multi-label classification presented in Table 4.9, the classification accuracy for the online validation is significantly lower. The difference is especially prominent for higher SNR levels. The results for WGN-contaminated EMG signals are more deficient than other contaminants.

The significant drop in classification accuracy suggests a mismatch between the training conditions and the online implementation. A key reason for this discrepancy may be that the models were not trained on sufficiently varied or representative data. For instance, what the model considers a clean signal may differ substantially from actual EMG activity during online use. The online validation was performed on the MyoCaptis device; no data from this device was used in model training. Aspects such as hardware, electrode placement, activity, and signal pre-processing could cause the signal to be significantly different from the training dataset. Moreover, the origin of the contaminants utilised for the online validation differs from those used for model training. WGN provides the most prominent example; the WGN was generated in MATLAB for training, while the WGN used for online validation was recorded. These two types of WGN differ greatly, as the WGN generated in MATLAB is a good approximation of white noise. In contrast, the recorded WGN is not purely white as it includes device-specific artefacts such as amplifier noise and electromagnetic interference. These deviations introduce structure into the signal, making it harder for the model to recognise it as WGN, decreasing the classification accuracy.

Two characteristics stand out from Table 4.10, Table 4.11, and Table 4.12: the performance gap between the single- and multi-label models, and the variation in accuracy between resting and moving segments. The performance gap between the single- and multi-label models in Chapter 5.5 is less pronounced in the online validation. The online evaluation is only performed on EMG signals with one contaminant present, removing the most challenging cases for the multi-label model, simplifying the task and likely improving the classification accuracy. The moving segments in the online validation tend to achieve higher accuracy than the resting segments,

likely due to the EMG amplitude increase during muscle activation, resulting in the added noise being scaled up to maintain the SNR level. As a result, the noise characteristics become more prominent, simplifying the detection of contaminants.

The models were not adapted to the user or device utilised for the online validation. The RF architecture can, to a certain extent, be fine-tuned by using a small amount of user-specific data. User-specific EMG characteristics, electrode placement, and hardware variability all influence the signal in subtle ways that the models may not account for. Incorporating such personalisation into the training process could improve accuracy and robustness in real-world use. The online validation generally performs better than the models presented in the literature; however, the trend of decreasing accuracy with increasing SNR is more pronounced in the online validation than in the offline validation. The trend is variably prominent for the different contaminations, supporting the mismatch theory between online and offline contaminants. Classification accuracy could be improved with fine-tuning and better representation of contaminants.

Another important factor influencing model performance is the frequency content of the training data. The EMG signals from public databases had already been bandpass filtered in the 20–500 Hz range, whereas the contaminant signals were largely unfiltered, introducing a mismatch in spectral characteristics. ECG and MA contaminants are mainly represented in the 0-20 Hz range, where non-contaminated EMG signals are mostly absent. As a result, the model may be overly sensitive to this range, potentially leading to biased learning and reduced robustness when applied to data where such noise is more subtle or removed.

An important consideration is the minimum SNR level required for a reliable operation of EMG devices, as this threshold directly influences their performance. Currently, the SNR required for a diagnostic tool or myoelectric controlled devices to operate accurately is unknown. Depending on the SNR requirements of those devices, the requirement for the contaminant classification models varies. If devices demonstrate robust performance at reduced SNR levels in natural conditions, the minimum SNR requirement for contaminant classification systems could be decreased from the 20 dB assumed in this project. Future research could focus on clarifying the SNR requirements for contaminant classification systems before further developing the classification system.

5.7 Future Work

The findings of this study provide a strong foundation for EMG contamination classification using machine learning with signal features, but there are also opportunities for further development. Future work could focus on enhancing model robustness and generalisability, exploring alternative model architectures, extending the approach to intramuscular EMG, and integrating the system into practical applications.

The provided model achieves a good classification accuracy when trained and evaluated online. For further improvement, developing a more robust and general model would require more data to be included; minimally filtered EMG data from additional sources could help generalise the model. When the model has a designated area of use, fine-tuning to the hardware and wearer should be performed to further enhance the classification accuracy. To provide a better model, additional contaminants should be included in the classification, allowing for a wider range of contaminants to be detected. Most notably, the discrepancy between offline and online performance points to a need for improved model generalisation. In this case, the contamination simulation should be reviewed to train and evaluate the model on realistic contaminants. Multiple methods are available to develop a more robust and generalistic model.

Diverting from the two presented models, a third model architecture could provide a better balance between classification accuracy and the ability to classify multiple contaminants simultaneously. A two-step model with an initial binary classification determining whether an EMG signal is contaminated or not, followed by a multi-label model classifying the eventual contaminations. This model would take advantage of strengths from the individual models, the speed and accuracy from the single-label model, and the ability to classify a signal with multiple contaminations from the multi-label model. Such a model could serve as a useful indicator for end users, such as prosthesis wearers or healthcare personnel performing diagnostic assessments, by detecting whether a signal is clean or contaminated.

Intramuscular EMG, recorded directly from muscle tissue, is less susceptible to surface-level artefacts, potentially simplifying classification. While intramuscular EMG is not immune to noise, its cleaner signal profile may improve model performance. Although this project focused on surface EMG due to its non-invasive nature, future work could explore how the current approach transfers to intramuscular EMG.

Finally, extensive validation in real-world conditions is essential. Moving beyond controlled offline and online testing, the system should be evaluated in clinical settings, such as diagnostic tools or prosthetic control. These contexts introduce more real-world variability in movement, electrode placement, user behaviour, and environmental interference; factors that are difficult to replicate in a lab setting. Testing in such conditions is necessary to assess the system's robustness, user-friendliness, and readiness for deployment outside of a research environment.

6
Conclusion

This thesis aimed to develop a machine learning model that offers information on potential contaminations in EMG signals. Based on the results presented throughout the thesis, it can be concluded that it is possible to develop a machine learning model offering such insight. With the seven effective and reliable signal features and two random forest models presented in the thesis, common EMG contaminants can be identified and classified in real-time under natural conditions. With the knowledge of occurring contaminations, better decisions can be made regarding filtering, which in turn would increase signal quality. This would allow devices dependent on EMG signals to operate more robustly, whether the application is of a diagnostic or assistive nature.

The results are based on systematic and statistical evaluations of signal features and machine learning models to identify the features that best differentiate between EMG signal contaminants and the model that provides the most accurate classifications. By systematically evaluating 41 signal features, seven features were selected for their ability to distinguish between different contaminations. The selected features, spanning both time and frequency domains, consistently differentiated clean from contaminated EMG signals across diverse datasets. Following tests of multiple classification models across a range of SNR levels, this thesis identified a Random Forest model as the most effective model architecture. Two models were developed: one tasked with single-label classification and one with multi-label classification. Generally, the models maintained strong classification performance even at high SNR levels, where contamination is less prominent, compared to current solutions, with the single-label model performing better than the multi-label model. These findings highlight the feasibility of building generalisable models for EMG contamination detection using five seconds of EMG data. Although online validation exposed challenges related to signal variability and generalisation, it also demonstrated the model's potential under realistic, dynamic conditions, laying the groundwork for further refinement and practical integration.

This work provides a foundation for EMG contamination detection using interpretable models and robust features. To move from proof-of-concept to real-world deployment, future work must focus on bridging the lab-to-field gap and increasing generalisability. This could be achieved through strategies such as fine-tuning models to individual users and hardware, incorporating more diverse and representative datasets, including minimally preprocessed EMG signals, and validating performance in more realistic settings. Doing so could accelerate the development

of reliable, user-friendly EMG-based systems for clinical diagnostics, prosthetic control, and other assistive technologies. Improving such technologies has the power to improve the user's quality of life.

Bibliography

[1] I. Karacan and K. S. Türker, "A comparison of electromyography techniques: Surface versus intramuscular recording," en, *European Journal of Applied Physiology*, vol. 125, no. 1, pp. 7–23, Jan. 2025, ISSN: 1439-6319, 1439-6327. DOI: 10.1007/s00421-024-05640-x. [Online]. Available: https://link.springer.com/10.1007/s00421-024-05640-x (visited on 01/29/2025).

[2] P. A. Parker and R. Merletti, Eds., *Electromyography: physiology, engineering, and noninvasive applications* (IEEE Press series in biomedical engineering), eng. Hoboken, N.J: Wiley-Interscience, 2004, ISBN: 9780471678380 9781601195098.

[3] M. Al-Ayyad, H. A. Owida, R. De Fazio, B. Al-Naami, and P. Visconti, "Electromyography Monitoring Systems in Rehabilitation: A Review of Clinical Applications, Wearable Devices and Signal Acquisition Methodologies," en, *Electronics*, vol. 12, no. 7, p. 1520, Jan. 2023, ISSN: 2079-9292. DOI: 10.3390/electronics12071520. [Online]. Available: https://www.mdpi.com/2079-9292/12/7/1520 (visited on 01/31/2025).

[4] D. Farina *et al.*, "The Extraction of Neural Information from the Surface EMG for the Control of Upper-Limb Prostheses: Emerging Avenues and Challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 797–809, Jul. 2014, ISSN: 1558-0210. DOI: 10.1109/TNSRE.2014.2305111. [Online]. Available: https://ieeexplore.ieee.org/document/6737308 (visited on 01/31/2025).

[5] J. Berning, G. E. Francisco, S.-H. Chang, B. J. Fregly, and M. K. O'Malley, "Myoelectric control and neuromusculoskeletal modeling: Complementary technologies for rehabilitation robotics," *Current Opinion in Biomedical Engineering*, vol. 19, p. 100 313, Sep. 2021, ISSN: 2468-4511. DOI: 10.1016/j.cobme.2021.100313. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468451121000532 (visited on 04/22/2025).

[6] E. Farago, D. MacIsaac, M. Suk, and A. D. C. Chan, "A Review of Techniques for Surface Electromyography Signal Quality Analysis," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 472–486, 2023, ISSN: 1941-1189. DOI: 10.1109/RBME.2022.3164797. [Online]. Available: https://ieeexplore.ieee.org/document/9749945/ (visited on 01/31/2025).

[7] J. Machado, A. Machado, and A. Balbinot, "Deep learning for surface electromyography artifact contamination type detection," *Biomedical Signal Processing and Control*, vol. 68, p. 102 752, Jul. 2021, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2021.102752. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421003499 (visited on 01/31/2025).

[8] A. Jena, P. Sharma, N. Gehlot, A. Vijayvargiya, and R. Kumar, "Efficient Contaminant Identification in sEMG Signals using Machine Learning," in *2024 Third International Conference on Power, Control and Computing Technologies (ICPC2T)*, Jan. 2024, pp. 25–30. DOI: 10.1109/ICPC2T60072.2024.10474881. [Online]. Available: https://ieeexplore.ieee.org/document/10474881 (visited on 01/31/2025).

[9] P. Phukpattaranont, N. Thiamchoo, and P. Neranon, "Real-time identification of noise type contaminated in surface electromyogram signals using efficient statistical features," *Medical Engineering & Physics*, vol. 131, p. 104 232, Sep. 2024, ISSN: 1350-4533. DOI: 10.1016/j.medengphy.2024.104232. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350453324001334 (visited on 01/31/2025).

[10] M. Irfan *et al.*, "Automatic Detection of Outliers in Multi-Channel EMG Signals Using MFCC and SVM," en, *Intelligent Automation & Soft Computing*, vol. 36, no. 1, pp. 169–181, 2022, ISSN: 1079-8587, 2326-005X. DOI: 10.32604/iasc.2023.032337. [Online]. Available: https://www.techscience.com/iasc/v36n1/50018 (visited on 02/12/2025).

[11] G. D. Fraser, A. D. C. Chan, J. R. Green, and D. T. MacIsaac, "Automated Biosignal Quality Analysis for Electromyography Using a One-Class Support Vector Machine," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2919–2930, Dec. 2014, ISSN: 1557-9662. DOI: 10.1109/TIM.2014.2317296. [Online]. Available: https://ieeexplore.ieee.org/document/6807760 (visited on 02/12/2025).

[12] M. Ait Yous, A. Said, and E. Siham, "A Fuzzy Inference System and Stationary Wavelet Decomposition for Identification and Removal of ECG Artifact from sEMG Signals," in *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, May 2024, pp. 1–7. DOI: 10.1109/IRASET60544.2024.10549577. [Online]. Available: https://ieeexplore.ieee.org/document/10549577 (visited on 02/12/2025).

[13] S. Thongpanja, A. Phinyomark, F. Quaine, Y. Laurillau, C. Limsakul, and P. Phukpattaranont, "Probability Density Functions of Stationary Surface EMG Signals in Noisy Environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 7, pp. 1547–1557, Jul. 2016, ISSN: 1557-9662. DOI: 10.1109/TIM.2016.2534378. [Online]. Available: https://ieeexplore.ieee.org/document/7438830 (visited on 02/12/2025).

[14] P. McCool, G. D. Fraser, A. D. C. Chan, L. Petropoulakis, and J. J. Soraghan, "Identification of Contaminant Type in Surface Electromyography (EMG) Signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 774–783, Jul. 2014, ISSN: 1534-4320, 1558-0210. DOI: 10.1109/TNSRE.2014.2299573. [Online]. Available: https://ieeexplore.ieee.org/document/6718122/ (visited on 02/12/2025).

[15] K. D. O. A. De Moura and A. Balbinot, "Virtual Sensor of Surface Electromyography in a New Extensive Fault-Tolerant Classification System," en, *Sensors*, vol. 18, no. 5, p. 1388, May 2018, ISSN: 1424-8220. DOI: 10.3390/s18051388. [Online]. Available: https://www.mdpi.com/1424-8220/18/5/1388 (visited on 02/12/2025).

[16] M. C. Tosin, L. B. Bagesteiro, and A. Balbinot, "Actor-Critic Reinforcement Learning Based Algorithm for Contaminant Type Identification in Surface Electromyography Data," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, ISSN: 2694-0604, Nov. 2021, pp. 186–189. DOI: 10.1109/EMBC46164.2021.9629967. [Online]. Available: https://ieeexplore.ieee.org/document/9629967 (visited on 02/12/2025).

[17] M. C. Tosin and A. Balbinot, "Identification and removal of contaminants in sEMG recordings through a methodology based on Fuzzy Inference and Actor-Critic Reinforcement learning," *Expert Systems with Applications*, vol. 206, p. 117 772, Nov. 2022, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2022.117772. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422010454 (visited on 02/12/2025).

[18] X. Zhang and H. Huang, "A real-time, practical sensor fault-tolerant module for robust EMG pattern recognition," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, p. 18, Feb. 2015, ISSN: 1743-0003. DOI: 10.1186/s12984-015-0011-y. [Online]. Available: https://doi.org/10.1186/s12984-015-0011-y (visited on 02/12/2025).

[19] G. D. Fraser, A. D. C. Chan, J. R. Green, and D. T. MacIsaac, "Biosignal quality analysis of surface EMG using a correlation coefficient test for normality," in *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2013, pp. 196–200. DOI: 10.1109/MeMeA.2013.6549735. [Online]. Available: https://ieeexplore.ieee.org/document/6549735 (visited on 02/12/2025).

[20] A. Ijaz and J. Choi, "Anomaly Detection of Electromyographic Signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 770–779, Apr. 2018, ISSN: 1534-4320, 1558-0210. DOI: 10.1109/TNSRE.2018.2813421. [Online]. Available: https://ieeexplore.ieee.org/document/8309308/ (visited on 02/12/2025).

[21] M. F. Issa, G. Tuboly, G. Kozmann, and Z. Juhasz, "Automatic ECG Artefact Removal from EEG Signals," en, *Measurement Science Review*, vol. 19, no. 3, pp. 101–108, Jun. 2019. DOI: 10.2478/msr-2019-0016. [Online]. Available: https://sciendo.com/article/10.2478/msr-2019-0016 (visited on 02/12/2025).

[22] M. K. Islam, P. Ghorbanzadeh, and A. Rastegarnia, "Probability mapping based artifact detection and removal from single-channel EEG signals for brain–computer interface applications," *Journal of Neuroscience Methods*, vol. 360, p. 109 249, Aug. 2021, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2021.109249. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027021001849 (visited on 02/12/2025).

[23] S. Phadikar, N. Sinha, R. Ghosh, and E. Ghaderpour, *Automatic Muscle Artifacts Identification and Removal from Single-Channel EEG Using Wavelet Transform with Meta-heuristically Optimized Non-local Means Filter*, arXiv:2201.01462, Apr. 2022. DOI: 10.48550/arXiv.2201.01462. [Online]. Available: http://arxiv.org/abs/2201.01462 (visited on 02/12/2025).

[24] S. I. Haider and M. Alhussein, "Detection and Classification of Baseline-Wander Noise in ECG Signals Using Discrete Wavelet Transform and Decision Tree Classifier," en, *Elektronika ir Elektrotechnika*, vol. 25, no. 4, pp. 47–57, Aug. 2019, ISSN: 2029-5731. DOI: 10.5755/j01.eie.25.4.23970. [Online]. Available: https://eejournal.ktu.lt/index.php/elt/article/view/23970 (visited on 02/12/2025).

[25] M. U. Abbasi, A. Rashad, G. Srivastava, and M. Tariq, "Multiple contaminant biosignal quality analysis for electrocardiography," *Biomedical Signal Processing and Control*, vol. 71, p. 103 127, Jan. 2022, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2021.103127. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421007242 (visited on 02/12/2025).

[26] D. Dharmaprani, H. K. Nguyen, T. W. Lewis, D. DeLosAngeles, J. O. Willoughby, and K. J. Pope, "A comparison of independent component analysis algorithms and measures to discriminate between EEG and artifact components," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, ISSN: 1558-4615, Aug. 2016, pp. 825–828. DOI: 10.1109/EMBC.2016.7590828. [Online]. Available: https://ieeexplore.ieee.org/document/7590828 (visited on 02/12/2025).

[27] J. Machado, M. C. Tosin, L. B. Bagesteiro, and A. Balbinot, "Recurrent Neural Network for Contaminant Type Detector in Surface Electromyography Signals," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, ISSN: 2694-0604, Jul. 2020, pp. 3759–3762. DOI: 10.1109/EMBC44109.2020.9175348. [Online]. Available: https://ieeexplore.ieee.org/document/9175348 (visited on 02/12/2025).

[28] M. Usman, M. Kamal, and M. Tariq, "Improved and Secured Electromyography in the Internet of Health Things," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2032–2040, May 2022, ISSN: 2168-2208. DOI: 10.1109/JBHI.2021.3118810. [Online]. Available: https://ieeexplore.ieee.org/document/9565394 (visited on 02/12/2025).

[29] W. R. Frontera and J. Ochala, "Skeletal Muscle: A Brief Review of Structure and Function," en, *Calcified Tissue International*, vol. 96, no. 3, pp. 183–195, Mar. 2015, ISSN: 0171-967X, 1432-0827. DOI: 10.1007/s00223-014-9915-y. [Online]. Available: http://link.springer.com/10.1007/s00223-014-9915-y (visited on 01/29/2025).

[30] C. McCuller, R. Jessu, and A. L. Callahan, "Physiology, Skeletal Muscle," eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK537139/ (visited on 01/29/2025).

[31] K. Mukund and S. Subramaniam, "Skeletal muscle: A review of molecular structure and function, in health and disease," en, *WIREs Systems Biology and Medicine*, vol. 12, no. 1, e1462, Jan. 2020, ISSN: 1939-5094, 1939-005X. DOI: 10.1002/wsbm.1462. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/10.1002/wsbm.1462 (visited on 01/31/2025).

[32] M. C. Gash, P. F. Kandle, I. V. Murray, and M. A. Varacallo, "Physiology, Muscle Contraction," eng, in *StatPearls*, Treasure Island (FL): StatPearls Pub-

lishing, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK537140/ (visited on 01/29/2025).

[33] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of NeuroEngineering and Rehabilitation*, vol. 7, no. 1, p. 21, May 2010, ISSN: 1743-0003. DOI: 10.1186/1743-0003-7-21. [Online]. Available: https://doi.org/10.1186/1743-0003-7-21 (visited on 03/21/2025).

[34] R. F. Ceballos and F. F. Largo, "On The Estimation of the Hurst Exponent Using Adjusted Rescaled Range Analysis, Detrended Fluctuation Analysis and Variance Time Plot: A Case of Exponential Distribution," 2018. DOI: 10.48550/ARXIV.1805.08931. [Online]. Available: https://arxiv.org/abs/1805.08931 (visited on 03/21/2025).

[35] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2012.01.102. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417412001200 (visited on 05/06/2025).

[36] *What Is Support Vector Machine? | IBM*, en, Dec. 2023. [Online]. Available: https://www.ibm.com/think/topics/support-vector-machine (visited on 03/07/2025).

[37] *What is the k-nearest neighbors algorithm? | IBM*, en, Oct. 2021. [Online]. Available: https://www.ibm.com/think/topics/knn (visited on 03/07/2025).

[38] *What Is Linear Discriminant Analysis? | IBM*, en, Nov. 2023. [Online]. Available: https://www.ibm.com/think/topics/linear-discriminant-analysis (visited on 03/07/2025).

[39] *9.2.8 - Quadratic Discriminant Analysis (QDA)*, publisher: Pennsylvania State University, 2018. [Online]. Available: https://online.stat.psu.edu/stat857/node/80/.

[40] *What is a Decision Tree? | IBM*, en, Nov. 2021. [Online]. Available: https://www.ibm.com/think/topics/decision-trees (visited on 03/07/2025).

[41] *What Is Random Forest? | IBM*, en, Oct. 2021. [Online]. Available: https://www.ibm.com/think/topics/random-forest (visited on 03/07/2025).

[42] *What Are Naïve Bayes Classifiers? | IBM*, en, Oct. 2021. [Online]. Available: https://www.ibm.com/think/topics/naive-bayes (visited on 03/07/2025).

[43] M. A. Ozdemir, D. H. Kisa, O. Guren, and A. Akan, "Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures," *Data in Brief*, vol. 41, p. 107 921, Apr. 2022, ISSN: 2352-3409. DOI: 10.1016/j.dib.2022.107921. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340922001330 (visited on 03/11/2025).

[44] S. Salter *et al.*, *Emg2pose: A Large and Diverse Benchmark for Surface Electromyographic Hand Pose Estimation*, arXiv:2412.02725 version: 1, Dec. 2024. DOI: 10.48550/arXiv.2412.02725. [Online]. Available: http://arxiv.org/abs/2412.02725 (visited on 03/11/2025).

[45] M. Atzori *et al.*, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," en, *Scientific Data*, vol. 1, no. 1, p. 140 053, Dec.

2014, ISSN: 2052-4463. DOI: 10.1038/sdata.2014.53. [Online]. Available: https://www.nature.com/articles/sdata201453 (visited on 03/11/2025).

[46] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 71, Jul. 2017, ISSN: 1743-0003. DOI: 10.1186/s12984-017-0284-4. [Online]. Available: https://doi.org/10.1186/s12984-017-0284-4 (visited on 03/11/2025).

[47] V. Behravan, N. E. Glover, R. Farry, P. Y. Chiang, and M. Shoaib, "Rate-adaptive compressed-sensing and sparsity variance of biomedical signals," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, ISSN: 2376-8894, Jun. 2015, pp. 1–6. DOI: 10.1109/BSN.2015.7299419. [Online]. Available: https://ieeexplore.ieee.org/document/7299419 (visited on 03/10/2025).

[48] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation [Online]*, vol. 101, no. 23, e215–e220, 2000. DOI: 10.1161/01.CIR.101.23.e215. [Online]. Available: https://doi.org/10.1161/01.CIR.101.23.e215.

[49] G. B. Moody, W. Muldrow, and R. G. Mark, *The MIT-BIH Noise Stress Test Database*, 1992. DOI: 10.13026/C2HS3T. [Online]. Available: https://physionet.org/content/nstdb/ (visited on 03/10/2025).

[50] M. Ait Yous, S. Agounad, and S. Elbaz, "Detection, identification and removing of artifacts from sEMG signals: Current studies and future challenges," *Computers in Biology and Medicine*, vol. 186, p. 109 651, Mar. 2025, ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2025.109651. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482525000010 (visited on 02/12/2025).

A

Appendix 1

Table A.1: Signal features from time domain

Signal Feature	References
Complexity Coefficient	[10]
Energy	[8]
Entropy	[22]
Hurst Exponent	[10]
Kurtosis 1, Kurtosis 2	[13]
Kurtosis	[8], [9], [12], [13], [22]
Mean Absolute Value	[8], [11], [15], [18]
Mean Frequency Length	[15]
Negentropy 1, 2, 3	[13]
Number of Slope Sign Changes	[18]
Number of Zero Crossings	[18]
Peak Amplitude	[8]
Peak-to-Peak Distance	[23]
Pearson Correlation Coefficient	[14]
Periodic Waveform Index	[22]
Power	[15]
Root Mean Square	[8], [10], [15]
Shannon Entropy	[23]
Signal-to-ECG Ratio	[14], [16]
Signal-to-Power Line Ratio	[14], [16]
Skewness	[8], [9], [12], [13], [22]
Ten-Bin EMG Histogram	[11]
Variance	[23]
Waveform Length	[15], [18]
Willison Amplitude	[11]

Table A.2: Signal features from frequency domain

Signal Feature	References
Anderson-Darling normality test	[25]
Composite multi-scale entropy	[12]
Maximum Drop in Power Density	[14], [16], [17]
Maximum Drop in Power-density Ratio	[8], [25]
Power Spectrum Deformation	[8], [16], [17]
Power Spectrum Deformation Ratio	[14], [25]
Shannon Entropy	[8]
Signal-to-ECG Ratio	[8], [17], [25]
Signal-to-Motion Artifact Ratio	[8], [14], [16], [17], [25]
Signal-to-Noise Ratio	[8], [14], [16], [17], [25]
Signal-to-Powerline Ratio	[8], [17], [25]
Spectral Entropy	[8]
Ten-Bin Averaged Power Spectral Density	[11]

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY