A Multiclassifier Approach to Motor Unit Potential Classification for EMG Signal Decomposition

by

Sarbast Rasheed

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Sarbast Rasheed

Abstract

EMG signal decomposition is the process of resolving a composite EMG signal into its constituent motor unit potential trains (classes) and it can be configured as a classification problem. An EMG signal detected by the tip of an inserted needle electrode is the superposition of the individual electrical contributions of the different motor units that are active, during a muscle contraction, and background interference.

This thesis addresses the process of EMG signal decomposition by developing an interactive classification system, which uses multiple classifier fusion techniques in order to achieve improved classification performance. The developed system combines heterogeneous sets of base classifier ensembles of different kinds and employs either a one level classifier fusion scheme or a hybrid classifier fusion approach.

The hybrid classifier fusion approach is applied as a two-stage combination process that uses a new aggregator module which consists of two combiners: the first at the abstract level of classifier fusion and the other at the measurement level of classifier fusion such that it uses both combiners in a complementary manner. Both combiners may be either data independent or the first combiner data independent and the second data dependent. For the purpose of experimentation, we used as first combiner the majority voting scheme, while we used as the second combiner one of the fixed combination rules behaving as a data independent combiner or the fuzzy integral with the λ-fuzzy measure as an implicit data dependent combiner.

Once the set of motor unit potential trains are generated by the classifier fusion system, the firing pattern consistency statistics for each train are calculated to detect classification errors in an adaptive fashion. This firing pattern analysis allows the algorithm to modify the threshold of assertion required for assignment of a motor unit potential classification individually for each train based on an expectation of erroneous assignments.

The classifier ensembles consist of a set of different versions of the Certainty classifier, a set of classifiers based on the nearest neighbour decision rule: the fuzzy k-NN and the

adaptive fuzzy k-NN classifiers, and a set of classifiers that use a correlation measure as an estimation of the degree of similarity between a pattern and a class template: the matched template filter classifiers and its adaptive counterpart. The base classifiers, besides being of different kinds, utilize different types of features and their performances were investigated using both real and simulated EMG signals of different complexities. The feature sets extracted include time-domain data, first- and second-order discrete derivative data, and wavelet-domain data.

Following the so-called *overproduce and choose* strategy to classifier ensemble combination, the developed system allows the construction of a large set of candidate base classifiers and then chooses, from the base classifiers pool, subsets of specified number of classifiers to form candidate classifier ensembles. The system then selects the classifier ensemble having the maximum degree of agreement by exploiting a diversity measure for designing classifier teams. The kappa statistic is used as the diversity measure to estimate the level of agreement between the base classifier outputs, i.e., to measure the degree of decision similarity between the base classifiers. This mechanism of choosing the team's classifiers based on assessing the classifier agreement throughout all the trains and the unassigned category is applied during the one level classifier fusion scheme and the first combiner in the hybrid classifier fusion approach. For the second combiner in the hybrid classifier fusion approach, we choose team classifiers also based on kappa statistics but by assessing the classifiers agreement only across the unassigned category and choose those base classifiers having the minimum agreement.

Performance of the developed classifier fusion system, in both of its variants, i.e., the one level scheme and the hybrid approach was evaluated using synthetic simulated signals of known properties and real signals and then compared it with the performance of the constituent base classifiers. Across the EMG signal data sets used, the hybrid approach had better average classification performance overall, specially in terms of reducing the number of classification errors.

Acknowledgements

In the Name of Allah, Most Gracious, Most Merciful.

All praise and thanks are due to Allah, and peace and blessings be upon His Messenger.

During the preparation of this thesis I have became indebted to many people for their patience, advice, and support. To all of them I owe very special thanks, and especially to my supervisors Professor Mohamed S. Kamel and Professor Daniel W. Stashuk.

I would like to thank my thesis readers and committee members: Professor George Freeman, Professor Otman Basir, Professor Eric Kubica, and my external examiner Professor Kevin Englehart for their stimulating conversations, intriguing ideas and suggestions.

Financial support for this research was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC).

I express my gratitude to my wife, my daughters, my sons, and to my family at my beloved city Baghdad - Iraq for their patience and support.

Dedication

This thesis is dedicated to my parents, my wife Nazdar,

my daughters Medea, Vian, and Maryam,

my sons Omar, and Mohammed,

my sisters and brothers.

Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

A_{FA}	Firing assertion decision function
A_i^j	MUP j overall assertion for belonging to MUPT i
A_m	Minimal assertion threshold
ACC	Adaptive Certainty Classifier
AFNNC	Adaptive Fuzzy k-NN Classifier
AMTFC	Adaptive Matched Template Filter Classifier
ANCCC	Adaptive Normalized Cross-Correlation Classifier Classifier
ApCC	Adaptive Pseudo Correlation Classifier
ANN	Artificial neural networks
$C(e_k)$	Confusion matrix of classifier e_k
C_{FC}	Firing certainty decision function
C_i^j	MUP j overall certainty for belonging to MUPT i
C_m	Minimal certainty threshold
CC	Certainty Classifier
Cf	Confidence associated with the classification of a MUP
CV	Coefficient of variation
d	Distance function, or
	Detail wavelet coefficients
$E(x)$	The overall decision of the classifier fusion system when input pattern is x
$e(x)$	Outcome of classifiers when input pattern is x
e_k	Classifier k
EMG	Electromyographic
EFE	Error-Filtered Estimation
F	Force, or
	Mapping function
$g^{i/k}$	Fuzzy density value of classifier e_k relative to class ω_i
$H(e_k)$	Confidence set of base classifiers set $\{e_k; k = 1, ..., K\}$
ID	Identification rate

IDI	Inter-Discharge Interval
K	Number of classifiers to be combined in an ensemble, or Number of modes in the Gaussian model implemented for the firing time confidence decision function
k	Positive integer representing the number of nearest neighbours
M	Number of classes, or Number of scale levels in wavelet decomposition
m	A MUP pattern feature vector, or Wavelet-domain scale index, or Fuzzification parameter used in the calculation of the class membership
MAD	Mean absolute deviation - the average of the absolute deviations of data points from their mean.
MFP	Muscle fibre potential
MTF	Matched template filter classifier
MU	Motor unit
MUP	Motor unit potential
MUPT	Motor unit potential train
MV	Majority voting classifier fusion scheme
MVC	Maximum voluntary contraction
N	Number of MUP patterns in an EMG signal, or MUP sample size
n	Number of MUP patterns in k-NN labelled reference set, or Wavelet-domain location index
NCC	Normalized cross correlation
NN	Nearest neighbour
P	Decision space, or Total number of classifiers in the pool of base classifiers
$P(\omega_i\|x)$	The a posteriori probability
pC	pseudo correlation
PDF	Probability density function
pps	The average number of MUP patterns per second
QEMG	Quantitative EMG analysis
s	A MUPT template vector, or A MUPT centre vector, or Approximation wavelet coefficients

S_i^j	MUP j overall similarity for belonging to MUPT i
S_{FS}	Firing similarity decision function
S_m	Minimal similarity threshold
SOFM	Self-organizing feature map
T_i	MUPT i template feature vector
t	Time
V	Set of labelled reference MUPs, i.e., the set of MUPs with known assignment
W_{ik}	Weight assigned to output i of classifier k
WT	Wavelet transform
X	Feature space
x	A MUP pattern feature vector, or Shifting position within the MUP pattern feature vector
$y^{(1)}$	MUP first-order discrete derivative feature vector
$y^{(2)}$	MUP second-order discrete derivative feature vector
α-motoneuron	Motoneuron that provides nerve supply to the muscle fibres located throughout the muscle
α	Majority voting fraction of the number of agreed classifiers
κ	kappa statistic for measuring the pairwise level of agreement
$\hat{\kappa}$	kappa statistic for measuring the level of agreement between many classifiers
μ	The MUPT mean IDI, or The class membership of a pattern
Ω	Set of class labels
ω_i	Train or class label
ρ	Matched template filter correlation
σ	The standard deviation of the distribution of MUPT IDIs

Chapter 1

Introduction

1.1 Preface

One approach to multiple classifier combination is classifier fusion, where classifier ensembles are applied concurrently and independently and an aggregator combines their results to achieve a group consensus.

Classifier fusion is based on combining different classifiers using different data representations, concepts and modelling techniques. The result for combining such a multimodal set of classifiers is different occurrences of classification errors for different classifiers over the set of patterns. The aggregation module of the classifier fusion system exploits such disagreement to errors of individual classifiers, and the greater this disagreement, the lower the effect of individual errors on the final decision, and effectively the lower the combined classification error.

This thesis addresses the process of EMG signal decomposition using a classifier fusion approach in order to achieve better classification performance. The proposed method jointly uses individual classifier decisions to make a final decision.

1.2 Electromyographic Signals

Electromyography is the detection of the electrical activity associated with muscle contraction. An electromyographic (EMG) signal is obtained by measurement of the electrical activity of a muscle during contraction, and reflects the electrical depolarization of excitable muscle fibre membranes that create electrical signals called muscle fibre potentials (MFPs).

The forward problem in electromyography is the composition of the electrical signal detected during a muscle contraction. In normal mammalian skeletal muscle, the fibres never contract as individuals. Instead, small groups of them contract in concert. All the fibres of each group of muscle fibres are controlled by the terminal branches of one nerve fibre or axon whose cell body is in the anterior horn of the spinal cord grey matter. These cells are the α-motoneurons and directly innervate skeletal muscle fibres.

α-motoneurons consisting of the nerve cell body, the long axon running down the motor nerve with its terminal branches, and all the muscle fibres controlled by these branches constitute a motor unit (MU). The summation of all of a MU's spatially and temporally dispersed MFPs results in a signal called a motor unit potential (MUP). In order to sustain a muscle contraction, motor units must be repeatedly activated and each motor unit generates multiple MUPs. The collection of MUPs generated by one motor unit, positioned at their times of occurrence or separated by their inter-discharge intervals (IDIs) is called a motor unit potential train (MUPT). The superposition of the MUPTs of all recruited motor units and background noise comprises an EMG signal.

Figure 1.1 shows a muscle structure and a schematic representation of the physiological model and instrumentation for the generation of an EMG signal. α-motoneurons in the spinal cord send electrical pulses to the muscle fibres, which cause the depolarization and contraction of groups of muscle fibres. The superposition at the detection site forms the physiological EMG signal $m_p(t, F)$ which is a function of time (t) and force (F). The integer p represents the total number of MUPTs which contribute to the potential field at

the detection site. When the signal is detected, an electrical noise, $n(t)$, is introduced. The detected signal will also be affected by the filtering properties of the detection electrode, $r(t)$, and possibly other instrumentation. The resulting signal, $m(t, F)$, is the observable EMG signal.

Clinically, EMG signal analysis, in the form of EMG signal decomposition and MUP classification into groups of similar shapes, is used to assist in the diagnosis of neuromuscular disorders, to analyze the neuromuscular system, and in biofeedback training.

The characteristics of an EMG signal are largely affected by anatomical and physiological properties of the muscle. For example, as the force of contraction increases, the number of motor units active and the rate at which they are active increases. The EMG signal therefore becomes more complex with increasing force of contraction. Furthermore, the fundamental structure of a muscle such as the size, distribution and number of motor units and how they are controlled can also be reflected in the characteristics of an EMG signal. For these reasons, many researchers are interested in devising techniques for the quantitative analysis of EMG signals.

1.3 EMG Signal Decomposition

The inverse problem in electromyography consists of using the detected EMG signal to infer the MUPTs of the recruited MUs comprising the EMG signal and perform EMG quantification. The process of resolving a composite EMG signal into its constituent MUPTs is called EMG signal decomposition. Figure 1.2 shows the separation of the MUPs of a segment of an EMG signal into ten superimposed MUPTs.

The objective of EMG signal decomposition is often the extraction of relevant clinical information from quantitative EMG (QEMG) analysis of individual MUPs and MU firing patterns. The first task in EMG signal decomposition is the segmentation of the EMG signal and detection of possible MUP waveforms, which is then followed by the main task

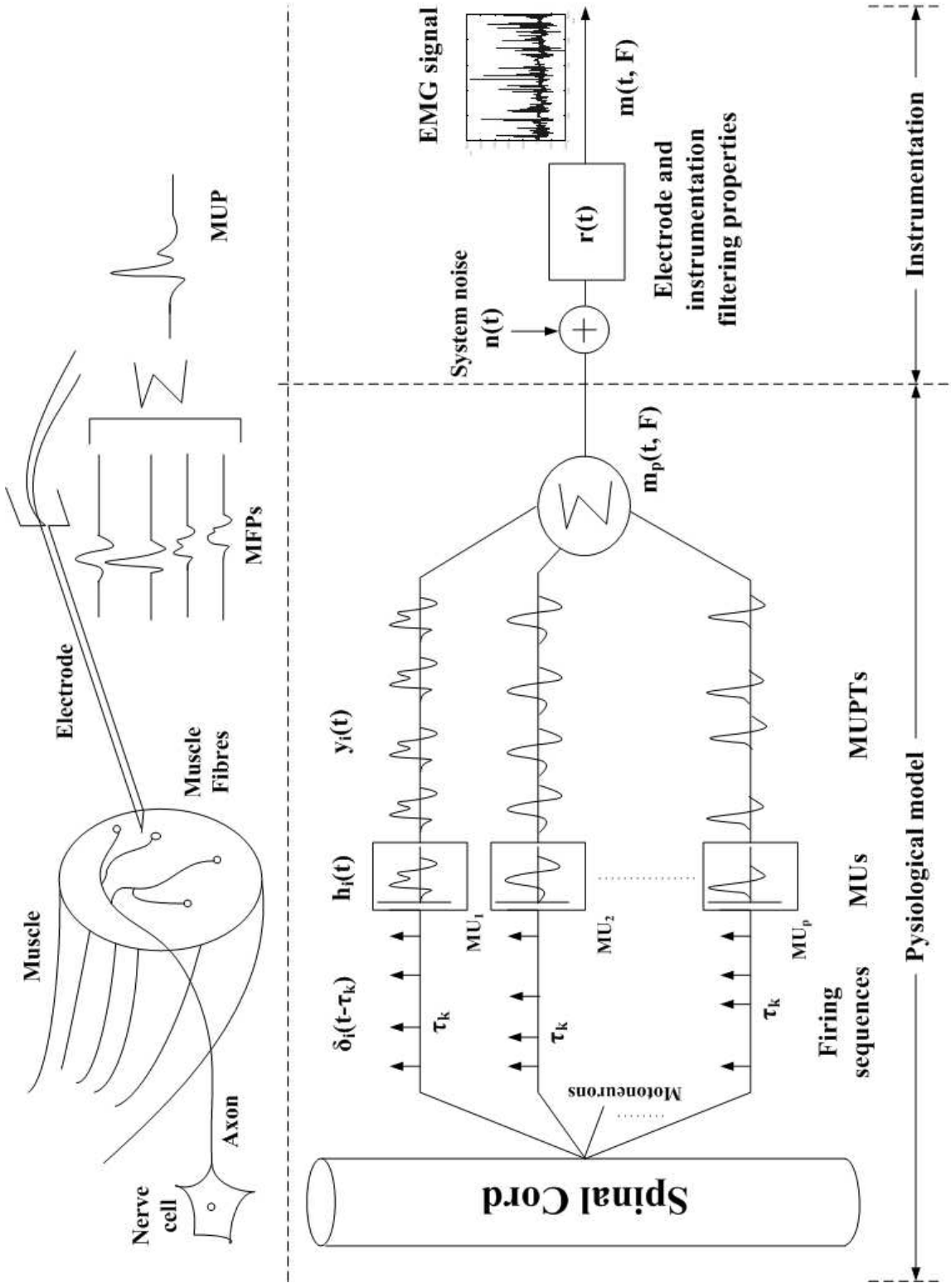

Figure 1.1: Schematic representation of the model for the composition of an EMG signal.

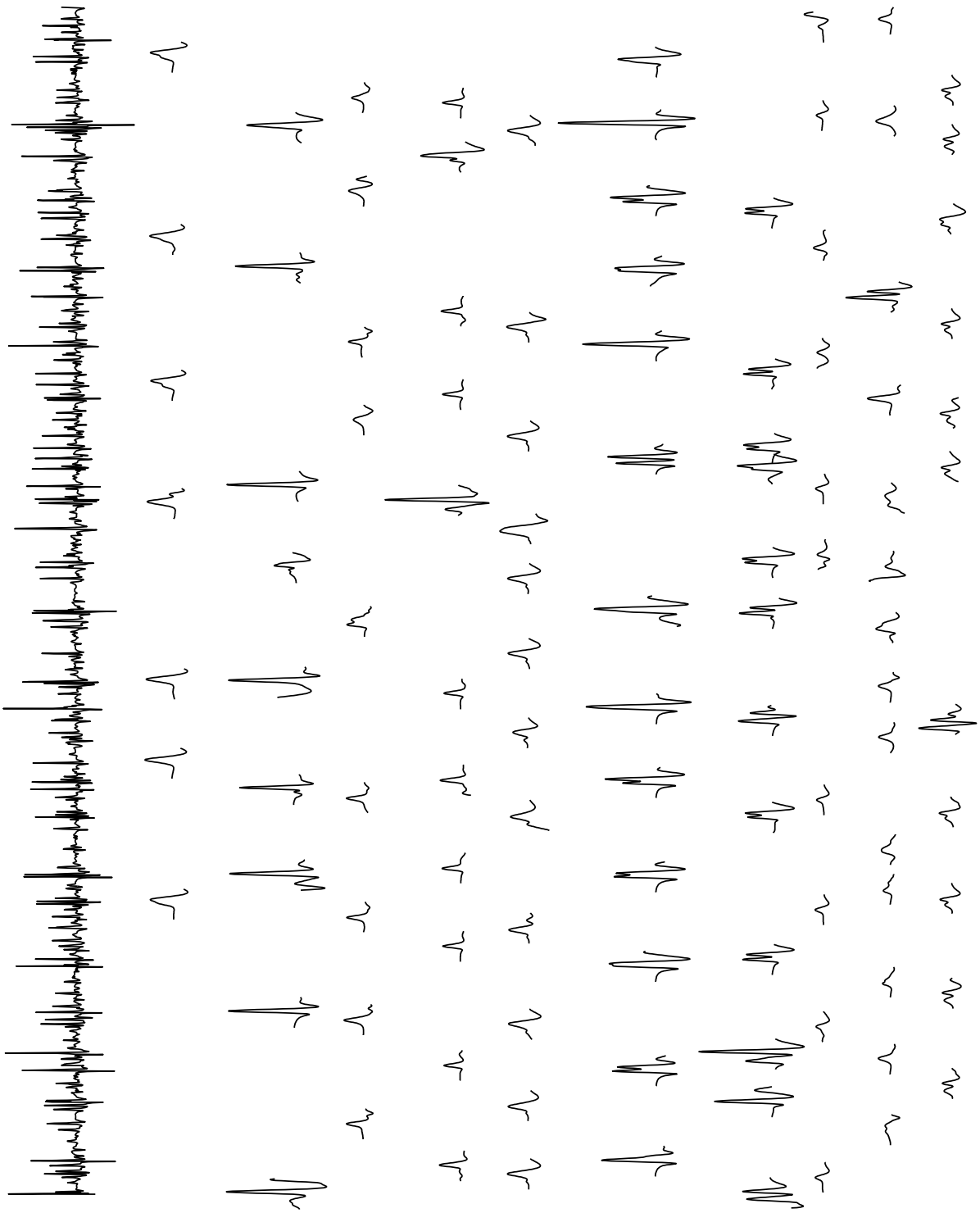

Figure 1.2: A segment of raw EMG signal decomposed into its constituent MUPTs.

of MUP classification. The classification task involves dividing detected MUPs into groups such that each set of grouped MUPs represents the discharges of a single MU and through which the discharges of each active MU can be discriminated for subsequent processing. QEMG analysis then often involves the calculation, for each MUPT, of a representative or template MUP waveform, which reflects information regarding individual MU morphology, and statistics related to the firing pattern of the MU.

The similarity criterion for grouping MUPs is usually based on a combination of MUP shape and statistics of the firing patterns of the motor units such that MUPs most likely belong to the same group if their shapes are closely similar and if their IDI interval is consistent with the discharge pattern of the considered motor unit. This means that two kinds of information, the MUP shapes and the times of occurrences of MUPs, should be considered for classification.

1.4 Combination of Multiple Classifiers

The combination of multiple classifiers can be considered as a generic pattern recognition problem in which the input consists of the results of individual classifiers, and the output is the combined decision [103]. It is based on the idea that classifiers with different methodologies or different features can complement each other. Hence if different classifiers cooperate with each other as a team, the combined decision may reduce errors drastically and achieve a higher performance.

Combining classifiers is now an established active research area and one of the main current directions in machine-learning research [19]. It has been applied to a wide range of real problems with the aim of overcoming the limitations of individual classifiers. In this thesis we are interested in applying this technique for the MUP classification task in EMG signal decomposition.

Combining multiple classifiers is called different names in the literature [56], [57]: com-

bination of multiple classifiers, classifier fusion, mixture of experts, committees of neural networks, consensus aggregation, voting pool of classifiers, dynamic classifier selection, composite classifier system, classifier ensembles, divide-and-conquer classifiers, modular learning, decision forest, and other names. The paradigm of these models differs in the assumptions about classifier dependencies, type of classifier outputs, aggregation strategy (local or global), and aggregation procedure (a function, a neural network, an algorithm).

In this thesis, we will use the term *classifier ensembles* to refer to the whole range of classifier combining methods and the term *classifier fusion* to refer to the situation when the individual classifiers are applied concurrently and independently.

1.4.1 Reasons for Combining Multiple Classifiers

The main motivation for combining classifiers is improving their performance, where multiple classifier systems try to exploit the local differences between base classifiers to enhance the accuracy and the reliability of the overall combined system. Combining a set of classifiers can be viewed as a way to manage the recognized limitations of the individual classifiers [111]. Each base classifier is known to make errors in such a way that the patterns that are misclassified by the different classifiers are not necessarily the same [55]. This means that the use of multiple classifiers can enhance the decision about the pattern to be classified.

There is also a situation in which some classifiers may be expected to fail in classifying some patterns and in this situation the overall combined system can recover the error.

1.4.2 Classifier Ensembles

Classifier ensembles are sets of base classifiers that work together to solve a recognition problem and whose decisions are combined to improve the performance of the overall system.

The methods that can be used to combine multiple classifier decisions depend on the type of information produced by the individual classifiers. Individual classifiers produce

information in any of the following three forms or a combination of them. Either a single class output that has the highest probability to which the input pattern belongs; or a ranked list of classes with the highest rank being the first choice; or a measurement value being assigned to each class label indicating the degree that the corresponding class pertains to the input pattern.

In this thesis we consider only combination methods based on the first and third types of classifier output information. Combination methods that can be applied when each classifier outputs a unique label or class for each input pattern consists of *voting schemes*. Combination methods that can be applied when each classifier outputs confidence values or certainty measures for each input pattern and for each target class, consist of methods that do not require prior training such as combination schemes based on the *product*, the *sum*, the *max*, the *min*, the *median*, and the *average* rules.

Instead of using one of the previous fixed combining rules, a training set can be used to adapt the combining classifier to the classification problem, where in this case the combination operator also functions as a classifier. The outputs of the base classifiers can be used as the input features of a general classifier used for fusion. The *fuzzy integral* method is one of the trainable combiners and is studied in this thesis.

1.4.3 Architecture of Multiple Classifier Systems

Combination strategies for classifier ensembles can be grouped into three main categories according to their architecture [47], [61], [82], [91]:

1. Parallel: In this case, all the individual classifiers are invoked concurrently and independently, and an aggregator combines their results.

2. Cascading (serial combination): In this case, individual classifiers are applied in succession one after the other, with each classifier producing a reduced set of possible classes for each pattern. Through this process, a complicated problem is progressively

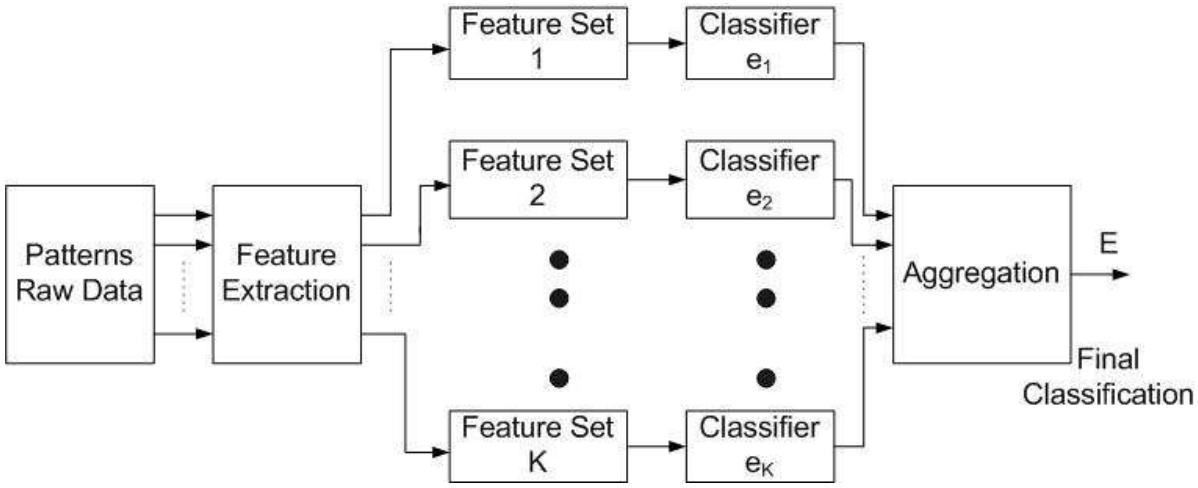

Figure 1.3: Classifier fusion system basic architecture.

reduced to simpler ones.

3. Hybrid: In this case, the advantages of both parallel and serial combinations are exploited to make the final decision more robust.

The combination architecture scheme implemented in this thesis belongs to the parallel category, whose basic architecture is shown in Figure 1.3, since our concentration is on classifier fusion methods, which require the individual classifiers be applied in parallel.

1.5 Objectives and Approach

The main objective of this thesis is to improve the performance and robustness of the classification task in EMG signal decomposition. We chose to implement a classifier combination paradigm such that we combine the decision of selected multiple classifiers with a goal to reach a combined decision with a higher performance, in terms of lower rejection rate and/or better accuracy rate, and get reduced performance variability across sets of EMG signals..

Traditional pattern recognition problems are solved mainly using a single feature identifier and a single classification procedure to assign a pattern to the class to which it should

belong. For difficult pattern recognition problems involving high dimensional patterns, large numbers of classes, and noisy inputs, perfect solutions are often difficult to obtain. However, improved solutions can be obtained by using systems of multiple classifiers as it is known that: usually decisions taken by teams are better than decisions taken by individuals. So that the objective of this thesis is to rely on a team's decision for providing better EMG signal decomposition through combining the decisions of different techniques that work as a team provided that suitable methods for choosing a team's base classifiers and combining the individual decisions are provided.

The developed research model is motivated by an attempt to achieve improved solutions for enhancing classification performance. It employs a hybrid classifier fusion approach for combining heterogeneous sets of classifier ensembles, which are classifiers built using different learning paradigms, for EMG signal decomposition.

1.6 Overview of the Thesis

This thesis explores a variety of classification paradigms for EMG signal decomposition beginning with individual base classifiers and ending with classifier combination techniques. Base classifiers include Certainty-based classifiers, nearest-neighbor-based classifiers, and matched template filter classifiers, and each with an adaptive and a non-adaptive version. Classifier combination techniques with different fusion schemes have been investigated.

We first review the EMG signal decomposition process and describe its modules. A survey of some of the existing methods is given. Then we review classifier ensemble techniques accompanied with a survey of existing methods. A multiple classifiers system architecture with fusion methods is given and a mathematical framework is derived.

The main contribution of this thesis begins from Chapter 3, where a set of base classifiers have been constructed and adapted for EMG signal decomposition.

In Chapter 4 we design a classifier fusion system for EMG signal decomposition. The

design involves two stages: the first stage deals with the design of the classifier ensemble, i.e., specifying which base classifiers comprise the team of classifiers for fusion, and the second stage deals with the design of the aggregation module. The developed classifier fusion model architecture is described including its base classifiers, model input space, aggregation module, and the need for data transformation when the model includes different types of classifiers.

Chapter 5 is devoted to results and comparative study. It presents the EMG signal data sets used consisting of two sets of simulated EMG signals, independent and related, and a set of real signals. The evaluation process for base classifiers and the classifier fusion system along with the approaches employed are described.

Finally, Chapter 6 summarizes the conclusions and presents some recommendations for future work to enhance the performance of the system.

Chapter 2

Background

2.1 Introduction

The objectives of this thesis are the use of multiple classifier fusion and the study of its effectiveness for EMG signal decomposition aiming to generate a more accurate classification than is possible from each of the constituent classifiers.

This chapter presents a survey of some of the existing partial and full EMG signal decomposition methods and a survey of multi-classification techniques with the related background necessary to build a multiple classifier fusion system.

2.2 Classification and Decomposition of EMG Signals

EMG signal decomposition is the process of resolving a composite EMG signal into its constituent MUPTs and it can be considered as a classification problem. Figure 2.1 shows the results of decomposing 1 s interval of an EMG signal, where the classifier assigns the MUPs into their MUPTs based on a similarity criterion. Those MUPs that do not satisfy the classifier similarity criterion are left unassigned.

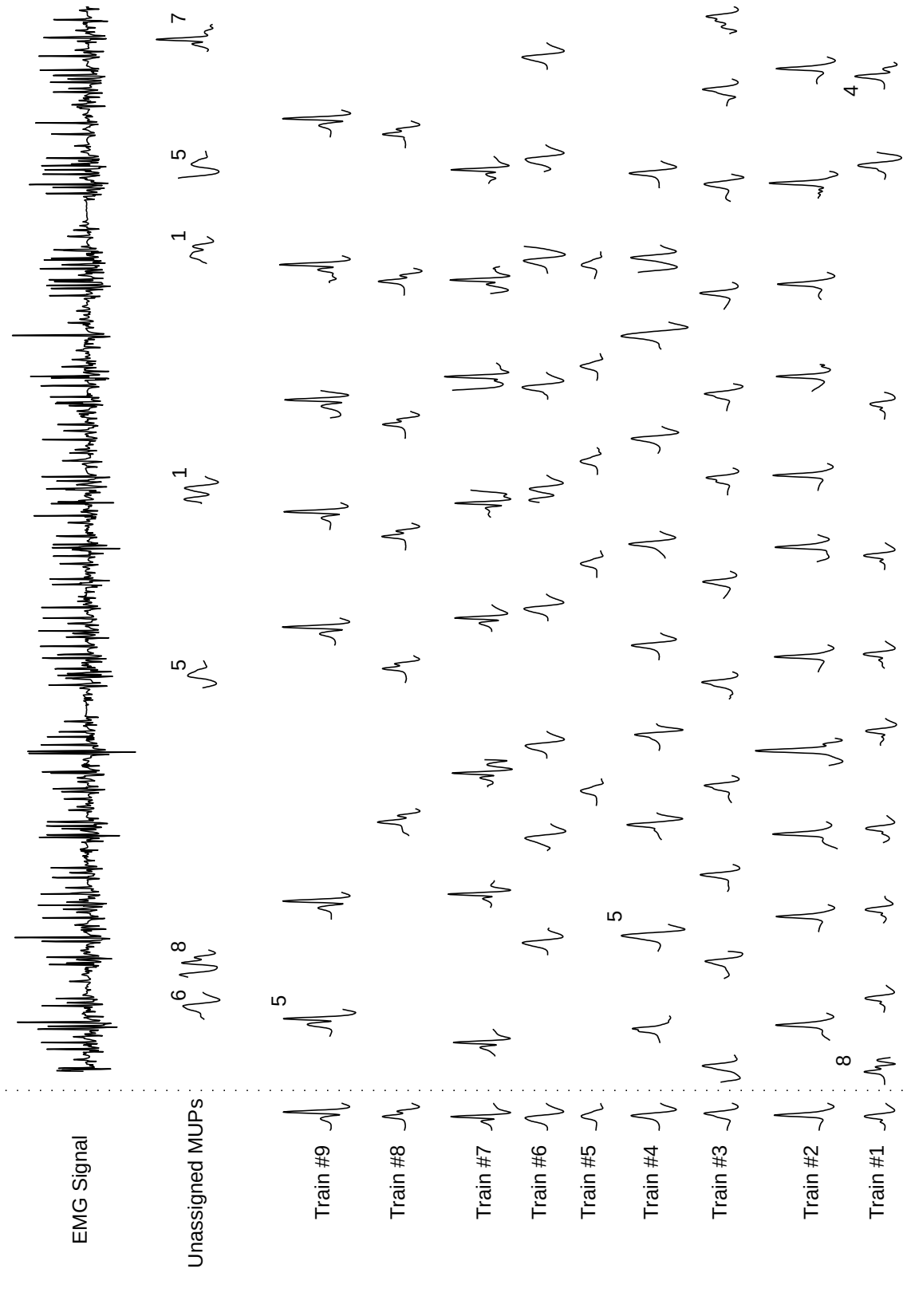

Figure 2.1: MUPTs for a 1 s interval of EMG signal decomposition. MUP waveforms have been expanded by a factor of 10 relative to the time scale used to depict their firing times.

Automatic EMG signals decomposition techniques have been designed to follow as closely as possible the manual method [3] and a good system should do the same analysis that an electromyographer does manually [26]. This is possible only if a robust pattern recognition algorithm is developed which is the purpose of this thesis.

Many automatic EMG signal decomposition techniques have been developed during the last two decades with different methodologies in the time, frequency, or wavelet domain being followed for quantitative analysis.

2.2.1 EMG Decomposition Process

A typical EMG signal decomposition system such as DQEMG (decomposition-based quantitative EMG) [96], [97] consists of the following modules, shown in Figure 2.2:

1. Signal acquisition and preprocessing,
2. Segmentation and MUP detection,
3. Clustering,
4. Supervised classification of detected MUPs,
5. Temporal relationships analysis,
6. Resolution of superimposed MUPs,
7. Quantitative EMG (QEMG): MUP characteristic features measurement and MU activation pattern analysis.

In this thesis we are mainly concerned with the supervised classification module.

An EMG signal acquired from a subject is fed to an analog preprocessing stage in which the signal is amplified and band-pass filtered (10 Hz to 10 kHz). The low frequency is set to 10 Hz to make the signal baseline more stable [95] by improving suppression of the baseline noise with minimum distortion to the MUP shapes [51]. The signal is sampled at

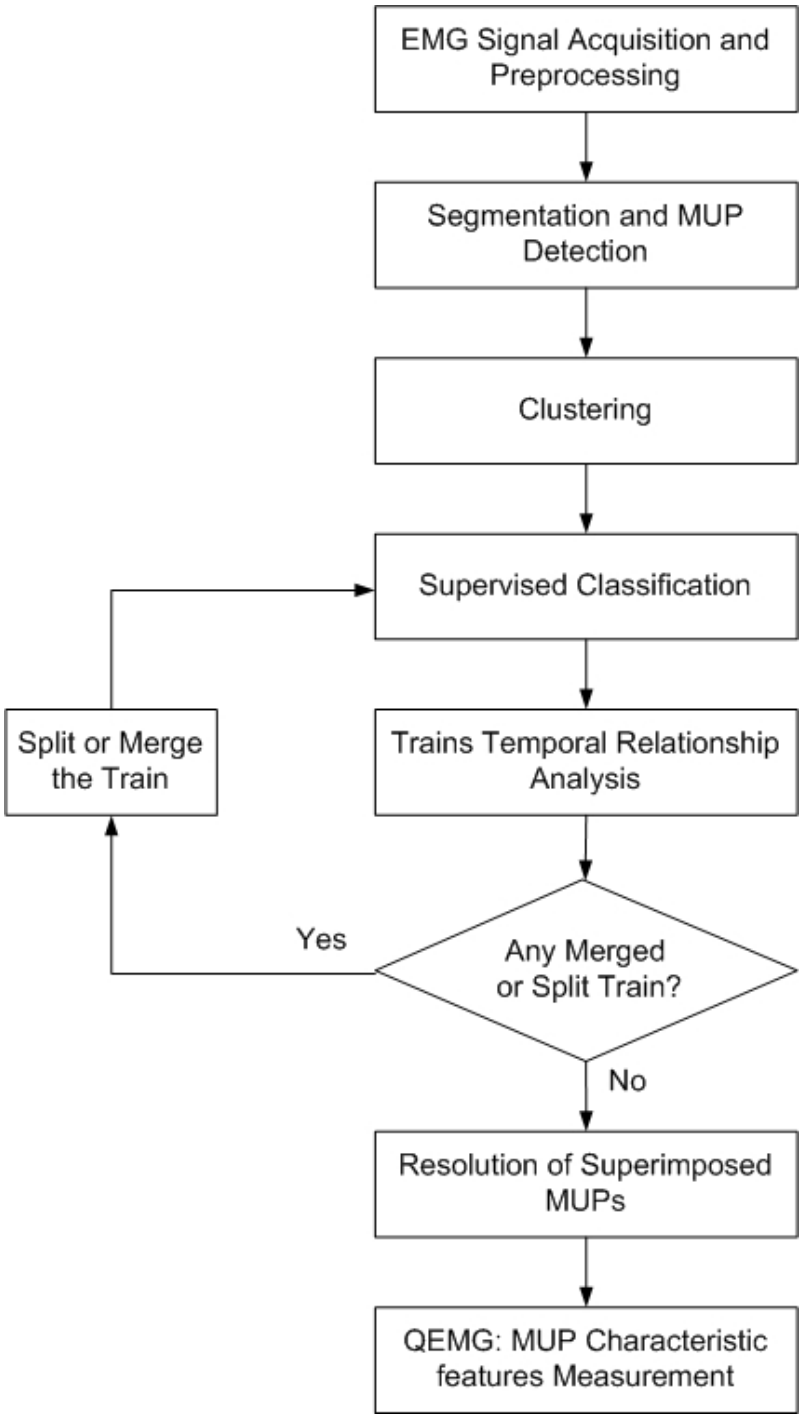

Figure 2.2: Schematic diagram of an EMG signal decomposition system.

a rate of 31.25 kHz and then digitized using a 12-bit resolution analog-to-digital converter. After acquisition, the signal is decomposed off-line.

The first task in EMG signal decomposition is the segmentation and MUP detection task where the signal is searched for time intervals containing MUPs. It is concerned with locating the main positive peaks or spikes found in an EMG signal. The detected spikes or MUPs should have rapid rising edges, which indicates that the electrode is close to active muscle fibres. Motor units that were active during signal acquisition generate these MUPs. Conversely, MUPs that have slow rising edges and small amplitude were generated by motor units whose fibres are far away from the electrode.

The EMG signal is divided into segments of possible MUP waveforms and searching for time intervals containing these MUPs defines the MUP detection operation. A segment can either contain one MUP or superimposed MUPs (compound segments). Time intervals with low energy are without MUPs and represent signal baseline. The detected spikes within windows of the sampled raw data or its first-order discrete derivative form the MUP waveforms. A window of 80 sample points represents MUP intervals of 2.56 ms and forms the MUP waveform.

In the case of real EMG signals, where there is no information about the number of recruited MUs that contributed to the signal, clustering (unsupervised classification) is an important step in EMG signal decomposition [100]. Therefore once a set of MUP waveforms have been detected and stored with their firing times, the part of the signal with the highest activity, i.e., the portion of the signal containing the largest temporal concentration of MUPs is chosen for clustering to provide the supervised classification stage with initial information about the number of active MUs (clusters) and the typical MUP shape or template for each motor unit. Once the number of active MUs and their templates are estimated, the classification stage is concerned with estimating the identities of all of the detected MUP waveforms.

The shapes and occurrence times of MUPs provide an important source of information to assist in the diagnosis of neuromuscular disorders. Some automatic EMG signal decomposition methods are designed so that the classification task considers only MUP morphological shape parameters such as: duration, amplitude, area, number of phases, number of turns, etc without evaluation of MU firing pattern or considering the variability of the MUP shape during contraction. These parameters can be used for diagnostic purposes since they reflect the structural and physiological changes of a MU. Others use MU firing patterns so that the central nervous system recruitment and control of MUs can be studied. Most of the new methods use both MUP shape parameters and either partial or full firing patterns [74].

The classification task for some of the existing decomposition methods are based on unsupervised classification, while others combine unsupervised and supervised classification methods. The unsupervised classification methods major limitation is that they only work well if there are large differences in the features of the classes involved [22] and because of the similarity between MUPs from different MUs, unsupervised classification methods often will not yield acceptable classification results [72]. Where they can result in lumping together two classes having similarly shaped MUPs into one class, or they can mistakenly separate one class into two classes [4]. On the other hand, a supervised classifier can track changing shapes over time, due to muscle fatigue and electrode or muscle movement, through updating the template of each train with each classification. This is done in the Certainty classifier [99] when tracking the non-stationarity of the MUP waveform shape, and in [116], [117] a weighted-average technique based on stochastic approximation is used to adapt the templates. Mirfakhraei et al. [72] used a bootstrap method for tracking the changing shapes through retraining a supervised artificial neural network classifier with new training sets generated based on the classification of the most recently acquired action potentials.

The classification task for EMG signal decomposition in this thesis was performed using multi-classification techniques through combining the results of a set of classifiers of different kinds and based on multi-features extracted from the acquired data. The classification scheme was based on information provided by the MUP waveform shapes and MU firing patterns.

2.2.2 Existing EMG Decomposition Methods

Most of the new automatic EMG signal decomposition methods use both MUP shape parameters and either partial or full firing patterns [74] for use during the classification task, which is based on either unsupervised classification or a combination of unsupervised and supervised classification methods. Below is a survey of some of the existing methods.

LeFever and De Luca [63] developed a decomposition technique that identifies and classifies MUPs based on both waveform template matching and firing time statistics. The method uses multiple channels of data to increase the power of identification during strong muscle contraction. The system is used to decompose EMG signals during constant force and force varying isometric contractions of up to 80% maximum voluntary contraction (MVC) with the capability of classifying up to eight concurrent MUPTs from the signal, and resolving MUPs formed by the superposition of up to two waveforms. The system is not totally automatic and requires operator intervention.

Stashuk and his research group [96], [97], [99], [100], [101] have developed an EMG signal decomposition system called DQEMG (decomposition-based quantitative EMG). DQEMG consists of a series of algorithms for estimating MU firing pattern statistics, clustering based on MUP shape and MU firing pattern characteristics, certainty-based supervised MUP classification, and determining temporal relationships between firing patterns of pairs of motor units.

Loudon et al. [64] used knowledge-based signal processing techniques for the decomposition of EMG signals. They developed a system that uses an expert-system architecture

and called DEMGES (decomposition of EMG expert system). The system automatically decomposes EMG signals recorded at force levels up to 20% MVC. Non-overlapping MUPs are classified using a statistical pattern recognition method. Whereas superimposed MUPs are decomposed using a combination of procedural and knowledge-based methods.

Hamid Nawab et al. [73] have also undertaken a knowledge-based approach to address some problems encountered during EMG signal decomposition. They showed that there is an increase in knowledge-based system's accuracy from 90% to well above 95% in decomposing complex EMG 3-channel data into its constituent MUPT. They concluded that the key to achieving this improvement is their use of a probabilistic framework for resolving pulse superpositions through the application of utility maximization at the suprasegmental level.

Gut and Moschytz [41] presented an approach to the decomposition of EMG signals based on a communication signal-based interpretation of the EMG signal. They analyzed the communication between neurons and muscles by communication technical means. The EMG signal source is modelled as an $(N + 1)$-ary, where N is the number of detected MUPs in the signal, digital signalling system with intersymbol-interference, which encodes a well-defined sparse information sequence. The information is conveyed by sending MUP waveforms from the source through a transmitter. The receiver observes the noisy EMG signal and has to estimate the information sequence, i.e., decodes the information sequence.

In recent years, artificial neural networks (ANN) have been used to classify MUPs. One of their major advantages is that ANN models make no assumption about the underlying probability density functions of the input data. Spitzer and Hassoun [94], and Hassoun et al. [44], [45] developed a system for EMG signal decomposition called NNERVE (neural network extraction of repetitive vectors for electromyography) capable of decomposing signals of up to 11 MUPTs. NNERVE utilizes an auto-associative algorithm, where the input vector serves as the target vector, and a pseudo-unsupervised learning approach using a customized error back-propagation algorithm for classification that uses the time domain

MUP waveform as input to a four-layer neural network. The input feature vector is 50 time domain samples and this is presented to the first layer, serving as an input layer, of the network consisting of 50 neurons. The second layer acts as a feature extractor and consists of 6, 8, or 10 neurons for different architectures. The third layer acts as an encoder by which the features from the second layer are encoded and consists of 4, 8, 12, or 16 neurons for different architectures. These codes are then fed to the output layer consisting of 50 neurons and in which the codes are transformed back into a set of features that represent the input MUP waveforms. The NNERVE method is computationally demanding due to the complicated architecture consisting of many layers, which require long training times and many learning epochs.

Instead of using time domain features, some decomposition methods employ other feature domains such as the frequency domain or the wavelet domain.

As a variation of the time domain concept of the matched filter technique, McGill [68], and McGill et al. [69] suggested a different way to perform template matching based on the coefficients of the Fourier transform. They developed a decomposition system called ADEMG (automatic decomposition electromyography) which can extract as many as 15 simultaneously active MUPs from EMG signals detected during contractions up to 30% MVC. ADEMG used template matching and a specific alignment algorithm for classification and achieved its high level of performance in identifying the MUPTs through several signal processing techniques. After identifying the trains, ADEMG examines their firing patterns to verify that they correspond to valid MUPTs. ADEMG was designed to be rapid and convenient for clinical use.

To reduce the dimensionality of the MUP patterns in the feature space, Stashuk and de Bruin [98] developed a method to decompose single fibre EMG signals using power spectrum matching. The power spectrum coefficients were used as features to represent the MUPs. Signals recorded during isometric, constant, or slow force varying contractions, up to 50% MVC were successfully analyzed with an accuracy of 95%. Since phase information

is lost with the calculation of the power spectrum coefficients, this affects the ability to discriminate between MUPs belonging to different MUs.

Wavelet features were used for classification in [27], [28]. They developed a technique to classify single MUPs and to decompose multiunit EMG signals based on spectrum matching in the wavelet domain and using the nearest neighbor clustering algorithm for classification. The developed technique was not suitable for on-line analysis and required operator intervention to reduce the decomposition error. Zennaro et al. [116] developed a method to decompose multichannel long-term intramuscular EMG signals in which the classification task is based on template matching using wavelet coefficients.

Fuzzy logic techniques have also been used for performing the classification task in EMG signal decomposition. Chauvet et al. [9] proposed an iterative algorithm with a classification method using fuzzy logic techniques to decompose an EMG signal detected during low to moderate force levels, where the number of detectable MUPTs is less than six. The fuzzy-based classification uses input variables derived from MU firing patterns and MUP information to refine the identification of each MUPT.

2.3 Classifier Ensemble Methods

Difficult pattern recognition problems involving high dimensional patterns, large numbers of classes, and noisy inputs can be solved efficiently using systems of multiple classifiers. Classifiers of different types, i.e., with different architectures, different classification procedures, and different feature spaces complement one another in classification performance and increase the probability that the errors of the individual features or classifiers may be compensated by the correct results of the rest. This has led to a belief that by using features and classifiers of different types simultaneously, classification accuracy can be improved [46] such that the performance of the classifier ensemble is never worse than the average of the individual classifiers, but not necessarily better than the best classifier [78].

There are a large number of methods, directions, and paradigms in designing classifier ensembles that have been proposed in the literature [115]. Ensemble techniques can be grouped and analyzed in different ways, depending on the main classification criterion adopted. Different categorization ways of classifier ensembles can be found in [50], [108], [109], [111].

2.3.1 Existing Classifier Ensemble Methods

To give an overview of the existing classifier ensemble methods, we follow the taxonomy proposed by Valentini and Masulli [108], [109]. They adopted an approach to distinguish between two classes of classifier ensembles: *generative* and *non-generative* methods considering the different ways base classifiers can be generated or combined together.

The class of generative classifier ensembles include methods that generate sets of base classifiers acting on the base classification algorithm or on the structure of the data set and try to improve the overall accuracy of the ensemble by directly boosting the accuracy and the diversity of the base classifiers. Below, we will describe some of these techniques:

1. Boosting: Is a general supervised method that is used to increase the accuracy of any classifier. The boosting algorithm originally developed was based on a theoretical model known as the *weak learning model* [90] and used to significantly reduce the error of any weak learning algorithm that consistently generates classifiers to achieve an arbitrary high performance. Several versions of boosting have been proposed, one of them is the AdaBoost (Adaptive Boosting) [33], and in particular AdaBoost.M1 algorithm, which can be used on classification problems with more than two classes. AdaBoost is an iterative algorithm with adaptive re-sampling of the training set in each iteration such that it generates classifiers sequentially. In each iteration it maintains a weight for each sample in the training set that reflects its importance, where it changes the weights of the samples based on the errors of previously generated classifiers. Adjusting the weights causes the classifier to focus on different examples

leading to different classifiers. For generating K base classifiers, AdaBoost in the beginning assigns all samples the same weight, thus forming a uniform distribution, and they are used to train the first classifier e_1. Then, the samples are re-weighted in such a way that the incorrectly classified samples have more weight than the correctly classified ones. Based on this new distribution, the classifier of the next iteration e_2 is trained. The classifier of iteration i, e_i, is therefore based on the distribution calculated in iteration $i-1$. Iteration by iteration, the weight of the samples that are correctly classified goes down and the weight of the incorrectly classified samples goes up. Therefore, the algorithm starts concentrating on the difficult samples. At the end of the procedure, K weighted training sets and K base classifiers have been generated. The final classification results from aggregating the base classifier results by weighted voting, i.e., for each sample, the output is a class label that maximizes the sum of the weights of the base classifiers predicting that label.

2. Bagging: Is an abbreviation of *bootstrap aggregating*. It can increase the classification accuracy significantly if the base classifier is properly selected. The bagging [7] algorithm is also not very sensitive to noise in the data. The algorithm uses the instability of its base classifier in order to improve the classification accuracy. Bootstrap methods are based on randomly and uniformly collecting samples with replacement from a sample set. The bagging algorithm is a supervised method and it constructs many different bags of samples by performing bootstrapping iteratively, classifying each bag, and computing some type of an average of the classifications of each sample via a vote. Bagging is in some ways similar to boosting, since both methods design a collection of classifiers and combine their conclusions with a vote. However, the methods are different. For example, because bagging always uses resampling instead of reweighting, it does not change the distribution of the samples (does not weight them), so all classifiers in the bagging algorithm have equal weights during the voting. It is also noteworthy that bagging can be done in parallel, i.e., it is possible to

prepare all the bags at once. On the other hand, boosting is always done in series, and each sample set is based on the latest weights. Each classifier e_i is trained on a bootstrapped set of samples from the original sample set. After all classifiers have been trained, a simple majority vote is used, but if more than one class jointly receives the maximum number of votes, then the winner is selected using some simple mechanism, e.g., random selection.

The class of non-generative classifier ensembles include methods that combine a set of base classifiers previously constructed with a suitable algorithms: they do not actively generate new base classifiers but try to combine in a suitable way a set of existing base classifiers.

In this thesis our goal is to combine a set of existing base classifiers, i.e., to combine a predetermined set of classifiers previously constructed. Based on that, our concentration here will be on non-generative ensemble methods.

There are two basic types of non-generative classifier ensemble methods [57], [114]:

1. Dynamic classifier selection: where each classifier is assumed as an expert in some local area of the feature space. The single classifier that is most likely to be correct for a given pattern is predicted. Only the output of the selected classifier is considered in the final decision.

2. Classifier fusion: where all classifiers are assumed to be expert across the whole feature space, and therefore their votes are equally important for any pattern. They are trained over the whole feature space and are considered competitive rather than complementary. Individual classifiers are applied in parallel and their outputs are combined in some manner to achieve a group consensus.

In this thesis we are interested only in classifier fusion methods, since for the dynamic classifier selection type, a method of partitioning the input patterns into selection regions

is required and then the best classifier for each partition is to be determined using training or validation data.

A large number of classifier fusion schemes have been proposed in the literature [115]. A brief description of some of the most popular methods will be presented.

1. Majority vote [62], [110]: In combining the decision of K classifiers, the pattern is assigned to the class for which there is a consensus, or when at least over half of the classifiers agree on a class label, i.e., if K is even ($\frac{K}{2}+1$) classifiers must agree or if K is odd ($\frac{K+1}{2}$) classifiers must agree. Otherwise, the pattern is rejected.

2. Simple averaging [110]: The individual classifier outputs are averaged across the ensemble of classifiers. The output yielding the maximum of the averaged values is chosen as the correct class:

$$E(x) = arg \max_{i=1}^{M} \left(\overline{e}_i(x) = \frac{1}{K} \sum_{k=1}^{K} e_{ik}(x) \right) \tag{2.1}$$

   where M is the number of classes, K is the number of classifiers, and $e_{ik}(x)$ represents the ith output of the kth classifier.

3. Weighted averaging [43]: Linear combination of the classifier outputs is performed with weights for individual classifiers. The combination-weights may be calculated according to any optimality criterion. Minimizing the mean squared error (MSE) over observed data may be used.

4. Bayesian combination [115]: The principle underlying this approach is to deal with the opinions of the classifiers as a data set. Then the aggregator combines the probability distributions provided by the classifiers via Bayes decision rule.

5. Borda count [46]: It is a generalization of the majority vote. The Borda count for a class is the sum of the number of classes ranked below it by each classifier and it measures the strength of agreement of the classifiers that the input pattern belongs

to that class. The combined ranking is given by arranging the classes so that their Borda counts are in descending order and the combiner output is the class with the largest Borda count.

6. Fuzzy logic based combination techniques [52], [105]: The fuzzy integral is a nonlinear functional that is defined with respect to a fuzzy measure. The fundamental basis of the fuzzy integral is that the priority of the processing or combination information is subjective to the degree of importance of the system, e.g., the most important system, such that the highest accuracy classifier contributes most to the final decision.

7. Stacked generalizations [113]: Is a general method of using a high-level model to combine lower level models to achieve greater predictive accuracy. Outputs from individual classifiers are combined in a weighted sum with weights that are based on the individual performance of the classifiers.

Recently, multiple classifier techniques have been used for classifying extracted MUP features for the purpose of assisting with the diagnosis of neuromuscular disorders.

Christodoulou et. al [15] developed a modular neural network system and showed its successful use in EMG decision making in mimicking the tasks carried out by an expert neurophysiologist in MUP analysis. The system employs a combination of multiple classifiers consisting of three types of neural networks: the back-propagation, radial basis function, and self-organizing feature map (SOFM) classifiers trained with time domain parameters, frequency domain parameters, autoregressive coefficients, cepstral coefficients, and four types of wavelet coefficients: the Daubechies 4, Daubechies 20, Chui, and Battle-Lemarie. The 24 output results of the eight features sets applied to the three classifiers were combined using majority voting. The system was trained and evaluated using five different bootstrap sets where in each set 24 different subjects were selected at random for training and 16 different subjects for evaluation. The mean and the standard deviation of the percentage of correct classification score, i.e., diagnostic yield, of the five different bootstrap

sets was computed for each classifier, for the eight different feature sets. It was found that the combination of different features and different classifiers based on the majority voting improved the overall classification performance of the system and gave a diagnostic yield higher than the average diagnostic yield of the individual feature sets.

Christodoulou and Pattichis [14] extended the above system so that the combination of different features with an ensemble of neural self organizing feature map (SOFM) classifiers can be done on a classifier combination measurement level using a confidence measure. The measure was derived from the SOFM classifier, which weighted the contribution of each feature set to the final classification result. The diagnostic yield obtained when combination was based on the confidence measure was improved further and it exceeded that obtained when combination was based on majority voting.

In a previous paper [84] we applied a classifier fusion approach for the classification task in EMG signal decomposition with a single type classifier ensemble consisting of six classifiers. The classifiers used were a modified version of the Certainty classifier [75], [97], [99] each fed with different feature set. Four classifier fusion schemes: majority voting, average fixed rule, Sugeno fuzzy integral, and Choquet fuzzy integral were applied separately and were used across a set of five simulated EMG signals, and all of them demonstrated improvement in the overall classification performance relative to the average performance of the individual classifiers across the five signals.

2.3.2 Fusion Methods

The next step after selecting the individual classifiers is combining the classifier outputs by a module, called the aggregator. Various aggregators can be distinguished from each other in their trainability, adaptivity, and individual classifier output requirements. Aggregators, such as voting, averaging or sum, are static, with no training required, while others require training. Some aggregators are adaptive in the sense that they evaluate (or weigh) the decisions of individual classifiers depending on the input pattern. In contrast, non-adaptive

aggregators treat all the input patterns the same. Adaptive combination schemes can further exploit the detailed error characteristics and expertise of individual classifiers [47].

Different classifiers typically express their decisions and provide information about identifying a pattern at different levels. Generally speaking, classifiers provide or are able to provide information divided into three levels [8], [115]:

1. Abstract level: the classifier output is a unique class label or several class labels, in which case the classes are equally identified without any qualifying information. A well known fusion method for this level is the *majority voting* [62], [110].

2. Rank level: the classifier output is a ranked list of the possible labels sorted by decreasing confidence without supplying the confidence. The *Borda count* [46] fusion method is the most frequently used rank level fusion approach.

3. Measurement level: the classifier attributes to each class label a confidence measure value representing the degree to which the pattern has that label. *Simple averaging* [110], *weighted averaging* [43], and *fuzzy integral* [52], [105] fusion methods are a few techniques for this level.

Different aggregators expect individual classifier outputs according to the aforementioned three output information levels. In this thesis, the developed system combines the individual classifiers outputs based on the abstract and measurement levels.

To combine the outputs from multiple disparate classifiers based on the abstract level we may use the voting strategy where the target class that receives the highest number of votes is selected as the final predicted class. Whereas, if the combination is based on the measurement level we may use methods that do not require prior training such as fixed combination schemes based on the product, sum, max, min, median, and average rules or use methods that require training.

The voting scheme neglects classifier differences in skills. This may be solved by assigning areas of expertise, following the best classifier for each new item of discussion, and,

in addition to the decision, the classifiers may be asked to provide some confidence [23].

To design an optimal decision procedure for combined classifiers, we need to evaluate the combination using simulated data with known class assignment, study the expert advice, and construct from that procedure a combined decision rule. In terms of classifiers this is called *training*.

2.3.3 Multiple Classifier Combination Framework

Xu et al [115] provided a mathematical framework for multiple classifier combination. To accomplish this, consider a decision space, P, with M mutually exclusive sets, $\omega_i \in \Omega = \{\omega_1, \omega_2, ..., \omega_M\}$. Each set, ω_i, represents a class or category into which patterns will be grouped or classified. The decision space may be written as:

$$P = \omega_1 \bigcup \omega_2 \bigcup ... \bigcup \omega_M \tag{2.2}$$

The decision space, P, is the set of all possible patterns from all classes. The set of the corresponding integer labels, Ω, is defined such that $\Omega = \{\omega_1 = 1, \omega_2 = 2, ..., \omega_M = M\}$ and it provides all possible integer labels for the defined classes. As some of the patterns may not be assigned to any of the available classes, the decision space set, P, can then be extended to include $\Omega \bigcup \{\omega_{M+1}\}$, where ω_{M+1} designates an unassigned class that by some established criteria the classifier has decided to not assign the input pattern.

For an ensemble of K classifiers, each recognition engine in the system may be simply regarded as a functional box that receives an input sample x and outputs a label, ω_j denoted by $e_k(x) = \omega_j$. This is regardless of what internal structure a classifier has or on what theory and methodology it is based. The team of K classifiers $e_1, e_2, ..., e_K$ provides a best choice when evaluating the input pattern x, in the form of an integer index, $w_j \in \Omega \bigcup \omega_{M+1}$ as a label indicating that x belongs to class ω_j.

Although the classification given by the index ω_j is the final product of any given single classifier, many existing classification systems can provide additional useful information.

For example, the Certainty classifier (CC) [75], [99] and the adaptive certainty classifier (ACC) [85] provide decision function values as a measure of certainty expressing confidence in the decision of classifying a MUP pattern to a particular train and the final label ω_j is the result of the maximum selection of the product of the decision function values. Once a classification has been made, the decision function values may not be retained; however, such discarded information may be useful for a multiple classifier fusion system when combining classifiers at the measurement level.

2.3.4 Abstract Level Combination

When classifying a pattern x at the abstract level, only the best choice is known from each classifier, $e_k(x)$. Therefore to combine abstract level classifiers, a voting method is used. The overall decision, $E(x)$, for the combined classifier system is sought given that the decision functions for the individual classifiers may not agree.

The most conservative form of voting is that all the individual classifiers must agree; otherwise the pattern is left unassigned. This requirement is a highly stringent condition and may lead to patterns remaining unassigned that might otherwise have been successfully classified.

A less conservative form than the above is that only all the classifiers that specify a preference need to agree, i.e., those choosing the unassigned category do not get a vote.

A more common and less stringent form of voting is the majority voting [62]. A pattern x is classified as ω_j if over half of the classifiers say $x \in \omega_j$. Majority voting can be generalized to the case in which some fraction of the classifiers specified by $0 < \alpha \leq 1$ are required to agree.

The voting rules expressed in a mathematical form are given in [115]. A generalized expression that includes all the aforementioned voting variants is given in (2.4). Representing the best choice from each classifier, $e_k(x)$, in the form of a binary characteristic

function:

$$T_k(x \in \omega_i) = \begin{cases} 1, & when\ e_k(x) = i\ and\ \omega_i\ \in\ \Omega \\ 0, & otherwise. \end{cases} \tag{2.3}$$

The general expression is:

$$E(x) = \begin{cases} \omega_j, & if\ T_E(x \in \omega_j) = max_{\omega_i \in \Omega}\ T_E(x \in \omega_i) \geq \alpha\ .\ K \\ \omega_{M+1}, & otherwise. \end{cases} \tag{2.4}$$

where ω_{M+1} designates the unassigned category.

2.3.5 Measurement Level Combination

When combining classifier outputs based on the measurement level we use the real valued outputs for each class provided by the respective classifiers. Assuming when classifying a pattern x, each classifier produces output values, in the interval [0, 1], interpreted as a confidence $Cf_i(x)$ in the decision of classifying pattern x with respect to a particular class ω_i $(i = 1, 2, ..., M)$. One can think of these outputs as a posteriori probabilities but it might be as a certainty measure about a class as is the case with the Certainty classifier [75], [99] or as an assertion measure about a class as is the case with the adaptive fuzzy k-NN classifier [83].

In terms of a posteriori probabilities, the confidence $Cf_i(x)$ is defined as:

$$Cf_i(x) = P(\omega_i|x) \tag{2.5}$$

and in relation with a specific classifier $e_k(x)$, the confidence depends on the outcome $e_{ik}(x)$ of this classifier for class ω_i:

$$Cf_{ik}(x) = P(\omega_i|e_{ik}(x)) \tag{2.6}$$

The confidences in (2.5) and (2.6) are defined over all the detected MUP patterns in an EMG signal.

Once a set of decision confidences $\{Cf_{ik}(x), i = 1, 2, ..., M; k = 1, 2, ..., K\}$ for M classes and K classifiers is computed for a MUP pattern x, they can be combined with a classifier fusion module into a new set of decision confidences that can be used, by maximum

selection, for the final classification. Either fixed combination rules or trainable combiners may be used.

In this thesis the *fuzzy integral* approach trained for a search for a set of densities was used for combining classifiers. The fuzzy integral [105] is a nonlinear, numeric-based approach for combining multiple sources of uncertain information. It uses a hierarchical network of evidence sources to arrive at a confidence value for a particular hypothesis or decision. In pattern recognition, the integral is evaluated over a set of features. Fuzzy integrals combine objective evidence for a hypothesis with the prior expectation of the importance of that evidence to the hypothesis.

A distinguishing characteristic of fuzzy integrals is that they utilize information concerning the worth or importance of subsets of information sources in the decision making process [52] and it is the only weighted aggregation operator, which takes into account not only the importance of information sources, but also the importance of all subsets of them [71]. The worth of subsets of information sources is represented by a fuzzy measure, a generalization of a probability measure. Fuzzy measures are not necessarily additive, which provides increased flexibility. Thus, fuzzy integrals can represent situations for which the measure of the whole is not equal to the sum of the parts.

Chapter 3

Base Classifiers

3.1 Introduction

Base classifiers are used in order to construct a combined classifier that may perform better than any of the base classifiers. Base classifiers should be different as it makes no sense to combine identical classifiers, but they should also be comparable, i.e., their outputs should be represented such that a combining method can use them as inputs. A consistent set of different classifiers may be generated in the following ways [23]:

1. Different initializations,
2. Different parameter choices,
3. Different architectures,
4. Different training sets,
5. Different feature sets.

The system developed in this thesis follows the so-called *overproduce and choose* [36], [76] strategy and the *test and select* [92] approach to ensemble combination by allowing the generation of a large set of candidate classifiers and then the choice of base classifiers team

is based on the selection of subsets of classifiers that can be combined to achieve better accuracy. The system employs, as base classifiers, different kinds of classifiers fed with different feature sets.

The base classifiers belong to three main types: certainty-based, fuzzy k-NN, and template matched filter classifiers. For each type there are different configurations. There are classifier configurations based on MUP shapes only, MUP shape and passive use of firing pattern information with an adaptive train-wise setting of assignment threshold, MUP shape and active use of firing pattern information with/without adaptive train-wise setting of the assignment threshold and others. The base classifiers similarity criterion is based on a combination of MUP shapes and a passive and an active use of MU firing patterns.

This chapter describes the passive and active use of MU firing patterns and presents the IDI statistics used for the detection of MUP misclassifications followed by a supervised classification of MUPs. Specifics of the Certainty, fuzzy k-NN, and matched template filter MUP base classifiers are described with the steps necessary to incorporate the IDI statistics in each of the mentioned base classifiers to form the adaptive version of each.

3.2 Motor Unit Firing Pattern Statistics

Two kinds of data are available for MUP classification in EMG signal decomposition: MUP shapes and the times of occurrences of MUPs such that the classification of MUPs to MUPTs cannot be considered independent of MU firing pattern constraints. These constraints include MU physiological limitations, the likelihood of a particular MU firing in a given epoch of time, and the expectation that the classification algorithm might not assign each detected MUP. For these reasons, the generated MUPTs may not represent the correct discharges of the active MUs and the need will arise to re-classify some MUPs.

Two modes of use of MU firing pattern information have been implemented: passive and

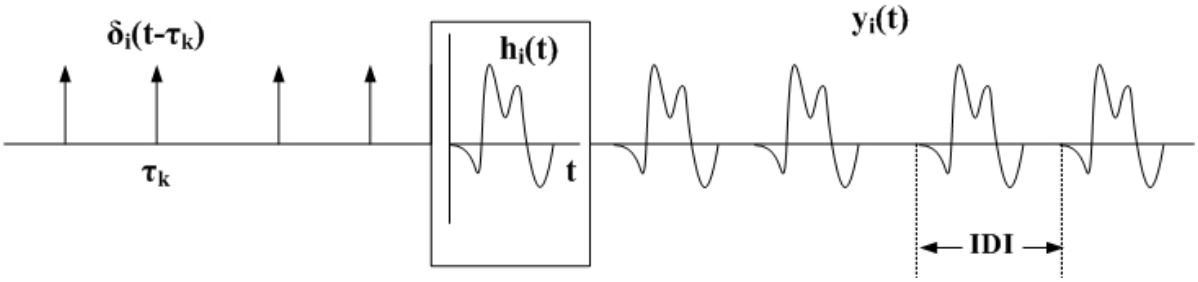

Figure 3.1: Model for a motor unit potential train.

active. Passive mode use refers to the use of firing pattern information in the adjustment of the assignment threshold of a MUP to a MUPT based on its firing pattern consistency statistics that adjust MUP assignments in general but not the assignment of any specific MUP. Whereas, active mode use refers to the situation in which firing pattern information is used to determine to which class each specific MUP should be assigned.

To take MU firing pattern constraints into consideration during a classification task, we specified a set of firing time consistency statistics that assist in the detection of MUP misclassifications.

3.2.1 IDI Statistics for the Detection of MUP Misclassifications

The MUPTs created by a classification algorithm can be modeled mathematically [5], [60], [65], by representing each train as a sequence of Dirac delta impulses, $\delta_i(t-\tau_k)$, with $k = 1, 2, ..., N_i$, where N_i is the number of discharges of MU_i and $i = 1, 2, ..., M$ where M is the number of MUPTs, which are passed through a filter whose impulse response is $h_i(t)$ as shown in Figure 3.1. If the impulses $\delta_i(t-\tau_k)$ mark the occurrence times of the train MUPs (i.e., MU firing times), the output of the filter would be the MUPT or $y_i(t)$ expressed by:

$$y_i(t) = \sum_{k=1}^{N_i} h_i(t-\tau_k) \tag{3.1}$$

It follows that a MUPT can be expressed as the convolution of the MU firing pattern

with a linear filter having an impulse response specified by the shape of the MUP. The MU discharge sequence can be described as a point process, which is a random process characterized by the times of occurrence of identical events.

Motor unit IDIs have been observed to be irregular and can be described as values of a random variable with characteristic statistical properties [65], and for finite samples of data, the IDI histogram serves as an estimator of the actual probability density function (PDF) [77]. Although Matthews [67] has demonstrated that the IDI PDF of a MU cannot actually be Gaussian, for MUPTs of MUs that are consistently recruited, the Gaussian density is a good approximation.

Given these expectations for the pattern of activity of a MU and the ability to characterize this pattern using the IDIs of a MU, errors in the determination of the activation pattern of a MU, caused by erroneous MUP classifications, can be detected by analyzing IDI statistics. Specific erroneous IDIs can be identified and determinations regarding the number of erroneous IDIs can be made. The following firing pattern consistency statistics can be used to determine if a significant number of IDI errors exist in a MUPT:

1. Percentage of inconsistent IDIs: an IDI inconsistency in a MUPT can be defined as those IDIs less than $\mu - 2\sigma$ and any IDI less than 15 ms. The latter constraint is related to the expected firing pattern of an α-motoneuron during constant force contractions. μ is the MUPT mean IDI and σ is the standard deviation of the distribution of MUPT IDIs. Using this definition of an inconsistent IDI, approximately 2.5% of the IDIs are expected to be inconsistent. If a greater percentage exists in a MUPT it can be assumed that a significant number of MUP classification errors have occurred.

2. IDIs coefficient of variation (CV) defined as:

$$CV = \frac{\sigma}{\mu} \tag{3.2}$$

Values of CV larger than 0.25 are suspicious since normal values are in the 0.1-0.15

range [5]. However, a large value of CV by itself does not necessarily represent MUP classification errors.

3. IDIs lower coefficient of variation (CV_l): defined as the ratio of lower standard deviation σ_l of the distribution of the MUPT IDIs and the MUPT mean IDI μ, and is given by:

$$CV_l = \frac{\sigma_l}{\mu} \tag{3.3}$$

σ_l is calculated using the IDIs whose values are less than the mean value. CV_l estimates the thickness of the lower portion (below the mean) of the IDI PDF. MUP classification errors tend to increase the proportion of shorter IDIs and increase CV_l accordingly. If there are no MUP classification errors, CV should approximately equal CV_l. The ratio of CV_l to CV measures the skewness of the IDI PDF to shorter IDIs and is a strong indicator of MUP classification errors.

4. Lower IDI ratio: defined as the ratio of the count of IDIs whose values are less than 0.5 of the MUPT mean IDI μ to the count of the IDIs whose values are less than the MUPT mean IDI μ. Like the CV_l and the ratio of CV_l to CV, the value of the lower IDI ratio will increase with the number of MUP classification errors. This statistic is however not dependent on estimates of σ or σ_l and therefore is not affected by errors in their estimation. This is especially important when a high value of σ and σ_l exists because in these situations it is difficult to get an accurate estimate of either.

To calculate the above firing pattern statistics, the mean μ and standard deviation σ of the train IDIs need to be calculated. The MUPTs identified during EMG signal decomposition may be incomplete since a train may have missing firings and they may also include erroneous firings. Therefore using all observed IDIs for estimating these firing pattern statistics does not work well even when the number of IDI errors is small. To get accurate estimates of μ and σ, with no a priori information about the firing pattern

of a MU and even when the number of IDI errors becomes large, an algorithm such as Error-Filtered Estimation (EFE) [81] must be used.

The EFE algorithm [81], [101] is used for estimating the mean and standard deviation of a set of time intervals between consecutive MU firing times (IDIs). It consists of three phases: in the first phase, the IDI histogram is constructed, and the set of IDIs is ordered such that the IDI histogram is divided into three regions whose boundaries are adjusted into: the region of low IDI values containing small IDIs due to false detection, the region of high IDI values containing large IDIs due to missed detection, and a region containing valid IDIs used to estimate the MU's true mean and standard deviation. Then the IDI means for each region is defined properly. In the second phase, the boundaries of the three regions are iteratively further adjusted so that the erroneous IDIs caused by missed detections and false detections are moved out of the region containing the valid IDIs into their corresponding regions. In the third phase, the final estimates are obtained based on the refinement done in the second phase.

As the input IDI data are filtered and only valid IDIs are used, the EFE algorithm provides accurate estimates even when the data defining the train MU firing times are only partially complete or have several erroneous firing times. The EFE algorithm has been found to provide accurate and unbiased μ and σ estimates, for moderate amounts of detection errors (up to 5%), even when up to 70% of the IDI data are missing [101].

To illustrate the behaviour of these firing pattern consistency statistics relative to MUP classification errors, a simulated MUPT of 200 discharges was generated with no errors and with a Gaussian distribution of the IDIs, whose mean IDI $\mu = 100$ ms and the IDIs distribution standard deviation $\sigma = 15$ ms such that its $CV = 0.15$. MUP false detections (equivalent to MUPs incorrectly classified) in the range of 0 - 10% and missed detections (equivalent to MUPs that are not classified) of 0, 20%, 30%, and 40% were consecutively introduced into the MUPT. For each MUPT configuration produced, the above four firing pattern statistics were calculated based on an EFE estimation of the mean IDI μ and

the IDIs distribution standard deviation σ. Figure 3.2 shows the behaviour of the firing pattern consistency statistics relative to the percentage of MUP classification errors for the different percentages of missed detections studied. This figure clearly shows how these statistics can be useful in detecting MUP classification errors, even for a small number of classification errors. Each of these statistics are individually correlated to the amount of false classification error and therefore correlated to each other. Nonetheless, because they have some conditional independence, we found that an increased number of false classification errors can be more consistently identified if all of these statistics are considered rather than using any of them individually.

3.3 Supervised Classification of MUPs

The task of supervised classification during the process of EMG signal decomposition is involved with the discrimination of the activation patterns of individual motor units, active during contraction, into distinguishable MUPTs. Therefore, MUPs are most likely belong to the same train if their shapes are closely similar and if their IDI intervals are consistent with the discharge pattern of the considered motor unit. This means that two kinds of information, the MUP shapes and the times of occurrences of MUPs, should be considered for classification.

For the purpose of MUP classification, we developed classifiers based on MUP shapes and with a passive and/or active use of firing pattern information. These classifiers follow an adaptive nature for train-wise setting of the assignment threshold based on firing pattern consistency statistics.

3.3.1 Adaptive Setting of Train Assignment Threshold

The adaptive aspect of the developed classifier is the setting of a minimal assignment threshold for each MUPT, which might be changed based on firing pattern statistics of the

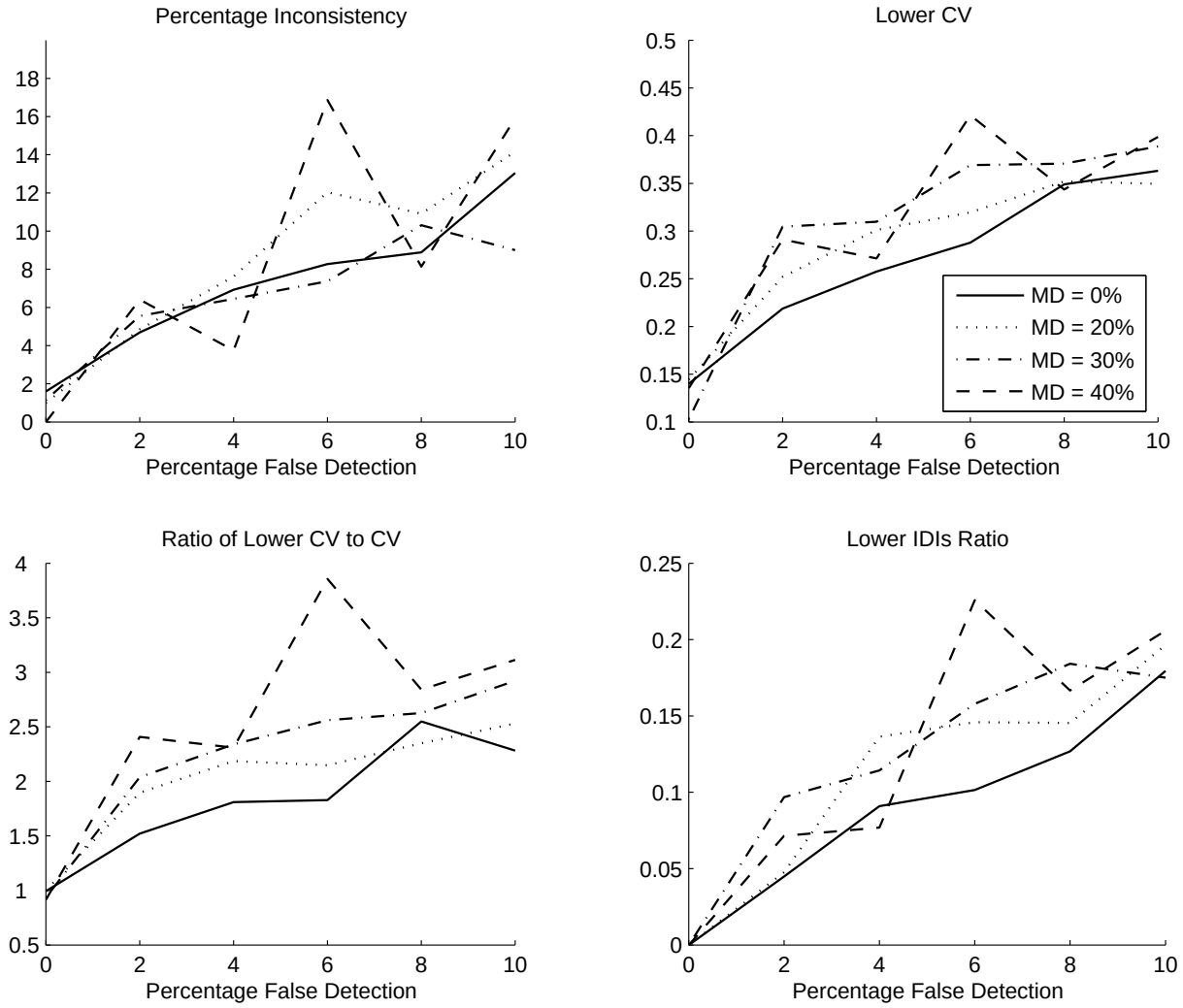

Figure 3.2: Firing pattern consistency statistics behaviour relative to false detection and with different percentages of missed detections. MD stands for missed detection.

train.

Following each classification pass through the MUP data, if, based on firing pattern statistics, it is expected that a train has too many erroneous assignments, its minimal assignment threshold is increased or otherwise it is decreased. This firing pattern analysis allows the algorithm to modify the required assignment of a MUP classification for each train individually based on an expectation of the number of erroneous assignments. Trains to which MUPs can be confidently assigned will have a lower minimal assignment threshold and have a higher MUP identification (ID) rate. Alternatively, trains to which MUPs cannot be confidently assigned will have a higher minimal assignment threshold and have a lower MUP identification rate. In both cases, the number of errors expected will be approximately constant. Therefore, over all the trains, a maximum number of MUPs should be assigned while maintaining an acceptable MUP assignment error rate. The adjustment of the minimal assignment threshold of each MUPT based on the pattern of MUP occurrences in the train represents a passive use of firing pattern information. Specific class assignments are not actively determined by the firing pattern information. Rather the firing pattern information is used to allow or deny MUP assignments in general.

During MUP classification, the classifier should not assign all detected MUPs as there are some detected MUPs considered superimposed MUPs comprising of more than one MUP or there are some MUPs causing dependent errors that have the effect of causing subsequent assignment errors when are assigned erroneously to a specific MUPT, so that the classifier should be able to recognize these MUPs and leaves them unassigned.

3.3.2 How does the Adaptive MUP Classification Work?

The adaptive nature of MUP classification is related to the adjustment of the minimal assignment threshold for each MUPT based on train firing pattern statistics. A MUPT assignment threshold might be increased to exclude MUPs causing firing pattern inconsistencies or decreased as long as firing pattern inconsistencies are not detected. The

occurrence of a significant number of MUP classification errors is detected by the use of the firing pattern consistency statistics presented in Section 3.2.1.

Due to the adaptive nature and the temporal dependence of the MUPs within a MUPT, the developed adaptive classifier employs an iterative multi-pass approach, where all of the MUPs, ordered based on their times of occurrence, are reconsidered for classification during each iteration. The iteration passes continue until specific stopping criteria are met.

The adaptive supervised classification task is divided into two stages:

1. The first stage involves only one classification pass that assigns MUPs based on shape only and passively uses firing pattern information to remove possible erroneous classifications. Figure 3.3 shows the steps involved in stage 1.

2. The second stage involves multiple classification passes and makes assignments based on both MUP shape and active and/or passive use of firing pattern information. Figure 3.4 shows the steps involved in stage 2.

Both classification stages are executed using an initial assignment threshold value. These values were empirically set based on experimentation applying the adaptive approach with different types of classifiers to several EMG signals such that the values used provided the best results or results not far from the best.

The classification passes during both stages are supervised classification passes. Each candidate MUP j is assigned to the train that has the highest confidence of belonging to it expressed in terms of the maximum assignment value and if it is higher than the minimum assignment threshold of that train. During the first stage, the highest confidence is measured in terms of the highest shape confidence of belonging to a train. Whereas, during the second stage, it is measured in terms of the best match of MUP shape and expected firing time represented by the multiplicative combination of the shape confidence value and the firing time confidence value.

Once each pass is ended, the firing pattern consistency statistics for each MUPT with

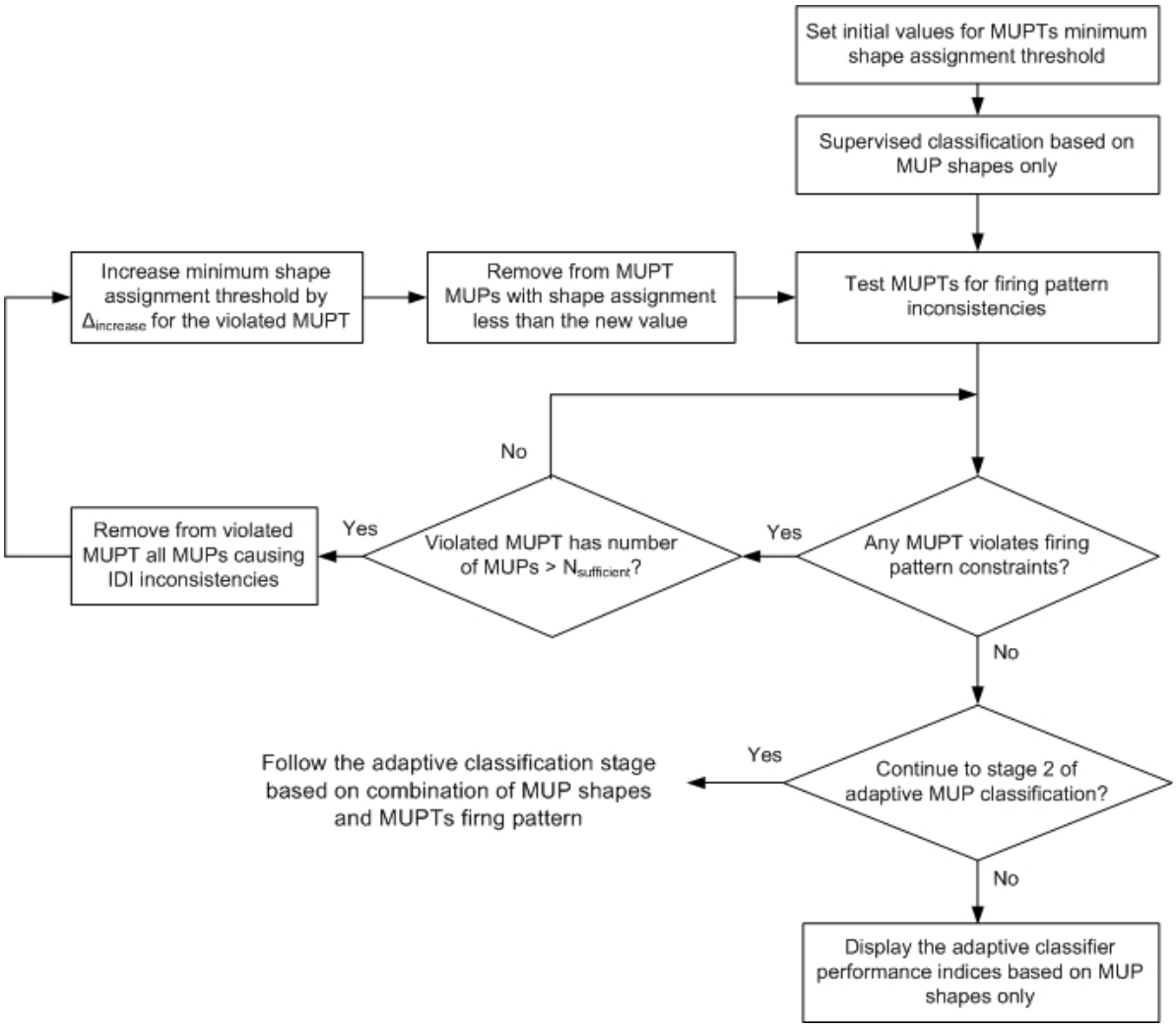

Figure 3.3: Flowchart of adaptive classification based on MUP shapes only.

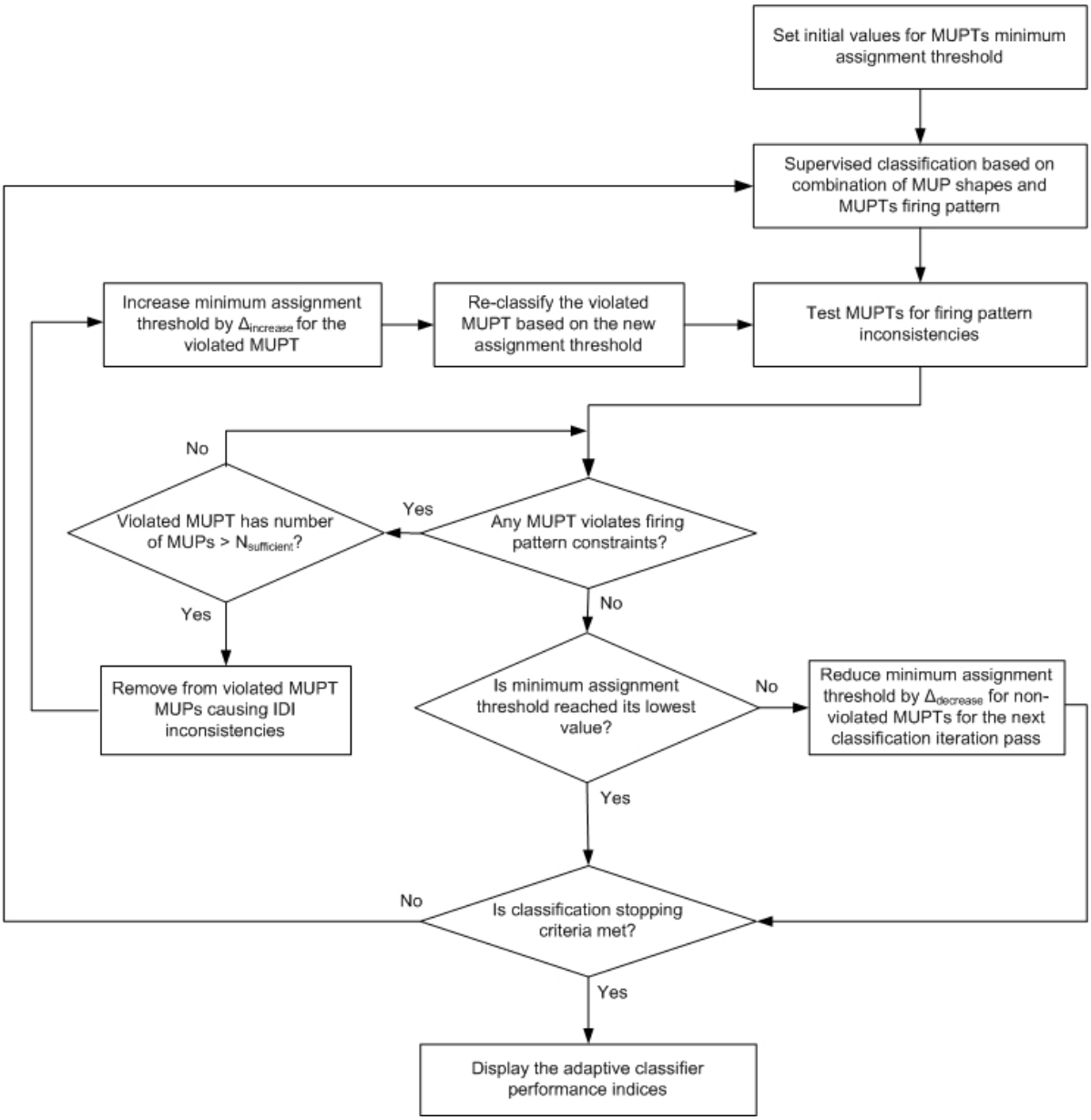

Figure 3.4: Flowchart of adaptive classification based on the combination of MUP shapes and MU firing patterns.

$N_{sufficient}$ or more IDIs are calculated. Then in the first stage all MUPs causing IDI inconsistencies in the firing pattern of a train are removed. While in the second stage, we allow a small percent of those MUPs causing IDI inconsistencies. The firing pattern consistency statistics of each MUPT with $N_{sufficient}$ or more IDIs are then re-calculated and checked against the following constraints:

Percentage number of IDI inconsistencies = 0% (in first stage) or 2.5% (in second stage).

Absolute value of $CV_l \leq 0.35$, representing an upper bound of physiological expectation.

Ratio of CV_l to $CV \leq 1.25$, empirically determined value.

Lower IDI ratio ≤ 0.175, empirically determined value.

If a MUPT meets all the above constraints, it keeps its MUPs and, if during the first stage, its minimum assignment threshold is unchanged while, if during the second stage, its minimum assignment threshold is reduced by $\Delta_{decrease}$. Otherwise, the minimum assignment threshold for the train is increased by $\Delta_{increase}$ and, if during the first stage, its MUPs confidence values are checked against the new value while, if during the second stage, its MUPs are re-classified using the new value of the assignment threshold. Those MUPs causing IDI inconsistencies in the firing pattern of a train and MUPs with confidence values less than the minimum assignment threshold are designated unassigned and removed from the train. This process is repeated until all the imposed firing pattern constraints for all generated MUPTs are satisfied. A MUPT's minimum assignment threshold is not decreased or increased below or above extreme values, respectively, or if it has fewer than a minimum number of MUPs $N_{sufficient}$. In all the experiments concerning the evaluation of base classifiers, the required number of MUPs in a MUPT to make minimum assignment threshold changes, $N_{sufficient}$, is taken to be 50 in order to enable the EFE algorithm estimates the MUPT mean and standard deviation with reasonable error.

The adjustment of the minimum assignment threshold of a train based on its firing pattern consistency statistic value at the end of the first shape-based pass and the subsequent shape and firing pattern passes, represents a passive utilization of MU firing pattern

information to adjust MUP assignments in general but not the specific class assignment of any specific MUP.

The classifier stops if no assignment threshold for a train is lowered and if no train has had a significant number of changes to its MUP assignments in the last pass, or if the maximum number of iterations (typically 10) has been completed.

3.4 Certainty-Based Classification

A certainty-based classifier classifies a candidate MUP to the MUPT that produces the greatest estimated certainty, provided this maximal certainty is above a minimum certainty threshold.

3.4.1 Certainty Classifier

The Certainty classifier (CC) is a template matching classifier that uses a certainty-based approach for assigning MUPs to trains. A complete description is given in [75], [97], [99] accompanied with testing and evaluation of its performance.

The CC estimates a measure of certainty expressing confidence in the decision of classifying a MUP to a particular train. It determines two types of decision functions for each candidate MUP, the first is based on shape information and the second is based on firing pattern information. For a set of MUPTs, the decision functions for MUP assignment are evaluated for only the two trains with the most similar templates.

The shape information decision functions include:

1. Normalized absolute shape certainty C_{ND}: represents the distance from a candidate MUP to the template of a train normalized by the norm of the template. For candidate MUP j, C_{ND1} is evaluated by:

$$C^j_{ND1} = max\{1 - \frac{r_1}{s_1}, 0\} \tag{3.4}$$

and C_{ND2} by:

$$C^j_{ND2} = max\{1 - \frac{r_2}{s_2}, 0\} \tag{3.5}$$

where, r_1 and r_2 are the Euclidean distances between MUP j and the closest (most similar) and second closest train templates, respectively; s_1 and s_2 are the l_2 norm of the closest and second closest train templates to MUP j, respectively.

2. Relative shape certainty C_{RD}: represents the distance from a candidate MUP to the template of the closest train relative to the distance from the same MUP to the second closest train. For candidate MUP j, C_{RD} is evaluated by:

$$C^j_{RD1} = 1 - \frac{r_1^2}{2\ .\ r_2^2} \tag{3.6}$$

$$C^j_{RD2} = \frac{r_1^2}{2\ .\ r_2^2} \tag{3.7}$$

where, C^j_{RDi} is the relative distance certainty associated with classifying MUP j to MUPT i.

The firing pattern information is represented by the firing certainty decision function with respect to the established firing pattern of the train. For candidate MUP j, C_{FC} is evaluated by:

$$\begin{aligned} C^j_{FC1} &= C_f(I^j_{b1}, \mu_1, \sigma_1)\ .\ C_f(I^j_{f1}, \mu_1, \sigma_1) \\ C^j_{FC2} &= C_f(I^j_{b2}, \mu_2, \sigma_2)\ .\ C_f(I^j_{f2}, \mu_2, \sigma_2) \end{aligned} \tag{3.8}$$

where, $C_f(I, \mu, \sigma)$ is a firing time certainty function based on the deviation of an IDI, I, from the estimated mean IDI, μ , of a train that has an estimated standard deviation, σ. In the current implementation of the CC, $C_f(I, \mu, \sigma)$ is evaluated using a multi-modal Gaussian model that takes into consideration missed-firings [68] given by:

$$C_f(I, \mu, \sigma) = \sum_{n=1}^{K} p_I^{(n)}(I) \tag{3.9}$$

where, $p_I^{(n)}(I)$ is based on a Gaussian probability density distribution:

$$p_I^{(n)}(I) \propto N(n\mu, n\sigma^2) \tag{3.10}$$

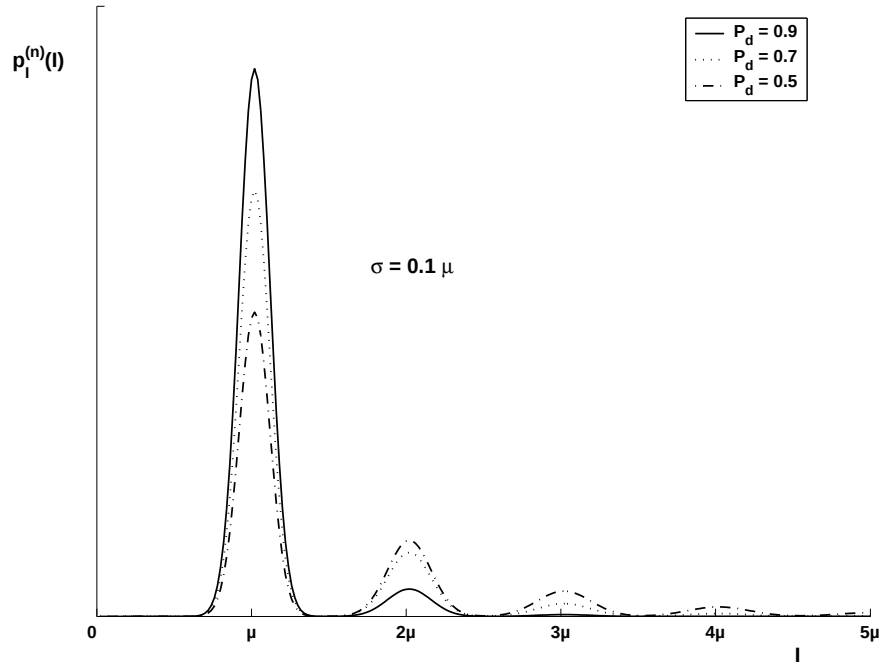

Figure 3.5: IDIs density distribution for MUPT with missing firings, mean $\mu = 50ms$, standard deviation $\sigma = 0.1\mu$, and detection probability $P_d = 0.5, 0.7, 0.9$.

$$p_I^{(n)}(I) = \frac{1}{\sqrt{n}} e^{-\frac{1}{2n\sigma^2}(I-n\mu)^2} \tag{3.11}$$

The modes become broader and smaller as n increases. In the current implementation K is set to a value of 40. Figures 3.5 and 3.6 show the firing time certainty function C_f for three MUPTs having detection probabilities of $P_d = 0.5, 0.7, 0.9$. P_d is assumed to be the same for all MUPs within a MUPT,

I_{bi} and I_{fi} are the IDIs that would be created by assigning a MUP j to train ω_i; I_{bi} is the backward IDI, the interval between MUP j and the previous MUP in the train; I_{fi} is the forward IDI, the interval between MUP j and the next MUP in the train. In (3.10) and (3.11), σ corresponds to the estimated standard deviation of the IDIs in the major mode ($n = 1$) from the major mode IDI mean μ.

The decision of assigning a MUP to a train is based on the value for which the multiplicative combination of C_{ND}, C_{RD}, and C_{FC} given by in (3.12) is the greatest and if it is

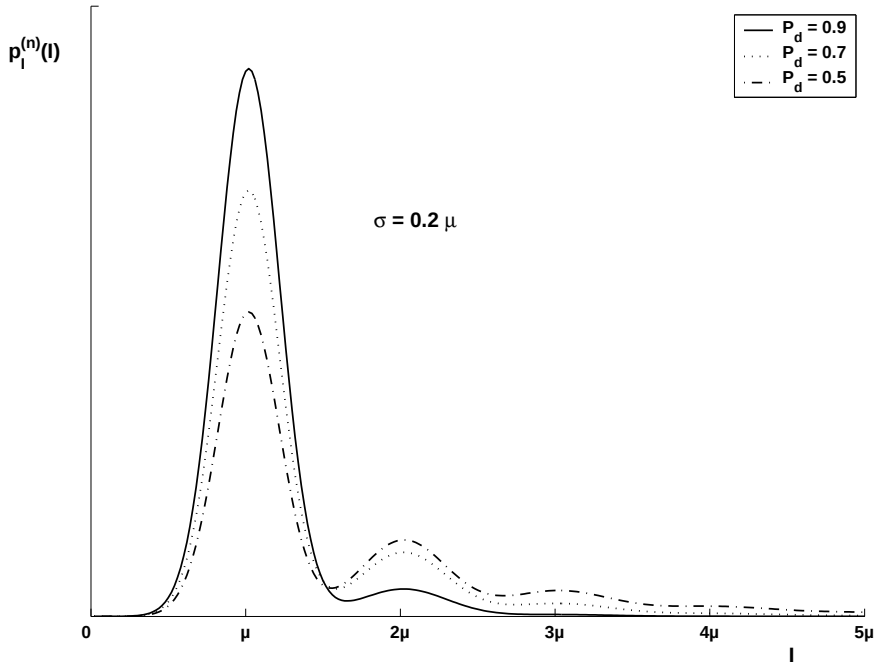

Figure 3.6: IDIs density distribution for MUPT with missing firings, mean $\mu = 50ms$, standard deviation $\sigma = 0.2\mu$, and detection probability $P_d = 0.5, 0.7, 0.9$.

greater than the minimal certainty threshold (C_m) for which a classification is to be made:

$$\begin{aligned} C_1^j &= C_{ND1}^j \cdot C_{RD1}^j \cdot C_{FC1}^j \\ C_2^j &= C_{ND2}^j \cdot C_{RD2}^j \cdot C_{FC2}^j \end{aligned} \tag{3.12}$$

where, C_i^j is the overall certainty associated with the classification of MUP j to MUPT i.

MUP j is assigned to the train with the closest template if

$$C_1^j > C_2^j \ and \ C_1^j > C_m \tag{3.13}$$

or to the train with the second closest template if

$$C_2^j > C_1^j \ and \ C_2^j > C_m \tag{3.14}$$

Otherwise, MUP j is left unassigned.

The value of each decision function is restricted to the interval [0, 1] and it corresponds to the confidence in the classification given the information of each function. A value of

1 corresponds to the ideal situation (maximum certainty) with respect to the information relevant to that function. However, for a classification that is certainly incorrect with respect to any source of information the corresponding decision function yields a value approaching 0.

3.4.2 Adaptive Certainty Classifier

The adaptive Certainty classifier (ACC) is a template matching classifier and it is a modified version of the Certainty classifier [75], [97], [99]. It uses an adaptive certainty-based approach for assigning MUPs to trains. The similarity criterion for grouping MUPs is based on a combination of MUP shapes and an active and passive use of MU firing patterns.

The adaptive nature of MUP classification is related to the adjustment of the minimal certainty threshold for each MUPT based on train firing pattern statistics and it follows the algorithms described in Section 3.3.2.

The classification pass during the first stage assigns each candidate MUP j to the train that has the highest shape certainty of belonging to it expressed in terms of the maximum shape certainty value calculated from the multiplicative combination of the maximum value of any of the two normalized absolute shape certainty given by (3.4) and (3.5) and the maximum value of any of the two relative shape certainty given by (3.6) and (3.7) such that it is assigned to the train given by:

$$\omega(j) = arg\ \max(C^j_{ND1}\ .\ C^j_{RD1},\ C^j_{ND2}\ .\ C^j_{RD2}) \tag{3.15}$$

if it is higher than the shape certainty threshold initially set to $C_m = 0.1$.

The classification passes in the second stage assign each candidate MUP j to the train that has the best match of MUP j shape and expected firing time only if this match has a certainty above the current certainty threshold for the train. The overall certainty of a MUP belonging to a train is given by (3.13) or (3.14) and measured in terms of the multiplicative combination of the normalized absolute shape certainty C_{ND}, the relative

shape certainty C_{RD}, and the firing time certainty C_{FC}. If the overall certainty is higher than the minimum certainty threshold initially set to $C_m = 0.02$, the candidate MUP j is assigned to the train given by:

$$\omega(j) = arg\ \max(C_1^j, C_2^j) \tag{3.16}$$

The adjustment of the minimal certainty threshold follows the adaptive setting of MUPT assignment threshold algorithm described in Section 3.3.2.

3.5 Nearest Neighbour Classification

The classification task in most EMG signal decomposition techniques is based on template matching. This section addresses the supervised MUP classification during EMG signal decomposition using the nearest neighbour (NN) classification rule instead of template matching. The primary use of nearest neighbour techniques involves situations where the a priori probabilities and class conditional densities are unknown [70]. Of the many nearest neighbour techniques, we chose the fuzzy k-NN classification rule, proposed by Keller et al. [53] as it deals with vagueness and uncertainty; and it provides a confidence measure regarding the classification results.

3.5.1 k-NN Classifier

The crisp nearest neighbour classification rule assigns a pattern, which is of unknown classification, to the class with the nearest neighbour. This idea has been extended to the k-nearest neighbours with the pattern being assigned to the class that has a majority of the k-nearest neighbours.

The k-NN classifier directly constructs the decision rule without explicitly estimating the class-conditional densities. The motivation of this classifier is that patterns which are close to each other in the feature space are likely to belong to the same pattern class [48].

The k-NN decision rule is a suboptimal procedure [22] and it provides a simple non-parametric procedure for the assignment of a class label to an input pattern based on the class labels represented by the k-closest neighbors of the input pattern [53]. The k-NN classifier does not rely on any assumption concerning the statistical distribution of data but instead it relies on a positive integer k, a metric or distance function between patterns d, and a labelled reference set that contains correctly classified patterns.

Let X be a feature space and let $\Omega = \{\omega_1, \omega_2, ..., \omega_M\}$ be the set of class labels, where M is the number of classes. Let $V = \{(v_1, l(v_1)), (v_2, l(v_2)), ..., (v_n, l(v_n))\}$ be the labelled reference set, where $v_i \in X$ and $l(v_i) \in \Omega$. An input pattern x of unknown classification is classified using the subset of k-labelled reference patterns that are closest to x with respect to the distance function d, i.e., pattern x is assigned to the majority class label corresponding to the k nearest neighbour. Statistically, this accounts for assigning the pattern x to the class with the highest a posteriori probability, i.e., the class most represented amongst the k nearest neighbors of x. The a posteriori probabilities are obtained as [56]:

$$P(\omega_i|x) \approx \frac{k_i}{k} \tag{3.17}$$

where k_i represents the number of neighbors belonging to class ω_i within the subset of k neighbors.

Figure 3.7 shows how a test pattern x is assigned to a class in a two-classes problem. The k-NN decision rule with $k = 5$ starts at the pattern x and grows a spherical region until it closes k nearest neighbors, and then it labels the pattern x by a majority vote of these neighbors. In this case, the pattern x would be labelled the class of black filled circles.

Two major problems are encountered when using a k-NN classifier. First, each of the neighbors is considered equally important in determining the classification of the input data. This frequently causes difficulty in those places where the sample sets overlap. A far neighbour to the input is given the same weight as a close neighbour. Second, the algorithm only assigns a class label to the input data, it does not determine the strength

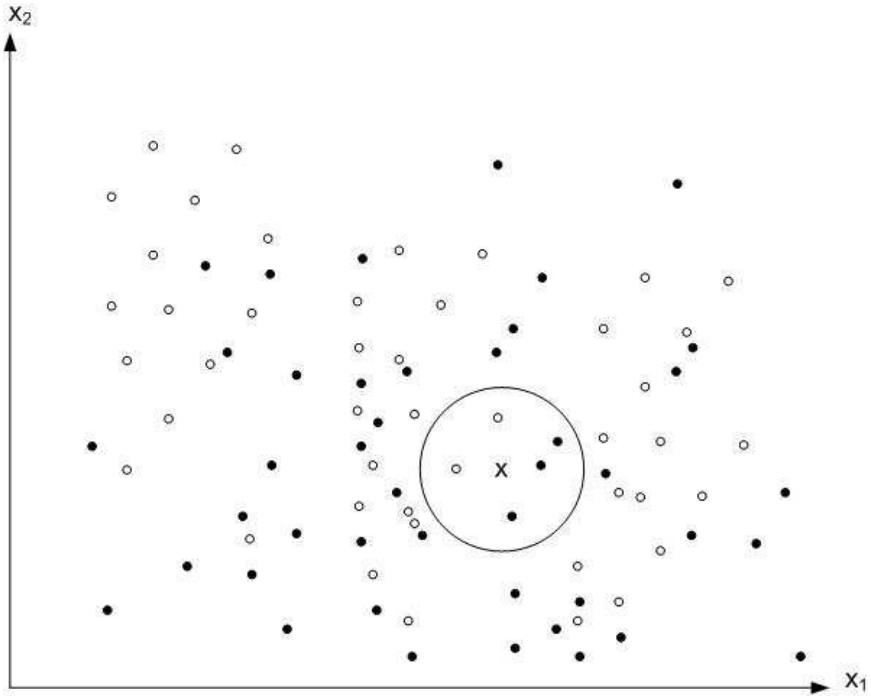

Figure 3.7: Application of k-NN decision rule to pattern x with $k = 5$.

of membership in the class. These two limitations can be addressed by the incorporation of fuzzy set theory into the k-NN rule and motivated the development of the fuzzy k-NN decision rule.

3.5.2 Fuzzy k-NN Classifier

The fuzzy k-NN classifier, designed by Keller et al. [53], is a fuzzy classification technique that generalizes the k-NN classifier. Rather than assigning a class label to an input patten x, the fuzzy k-NN algorithm assigns to pattern x a membership vector $(\mu_{\omega_1}(x), \mu_{\omega_2}(x), ..., \mu_{\omega_M}(x))$ as a function of the pattern's distance from its k nearest neighbors. This ensures that no arbitrary assignments are made. The class membership of the input patten x is calculated based on the following formula:

$$\mu_{\omega_i}(x) = \frac{\sum_{j=1}^{k} \mu_{\omega_i}(x_j)\ d_j^{-\frac{2}{(m-1)}}}{\sum_{j=1}^{k} d_j^{-\frac{2}{(m-1)}}} \tag{3.18}$$

where $x_1, x_2, ..., x_k$ denote the k nearest neighbour labelled reference patterns of x, and $d_j = \|x - x_j\|$ is the distance between x and its jth nearest neighbour x_j. The pattern x is assigned to the class given by:

$$\omega(x) = arg \max_{i=1}^{M}(\mu_{\omega_i}(x)) \tag{3.19}$$

The fuzzification parameter m determines how heavily the distance is weighted when calculating the class membership. As m increases toward infinity, the term $d_j^{-\frac{2}{(m-1)}}$ approaches one regardless of the distance. Consequently, the neighbors x_j are more evenly weighted. As m decreases toward 1, however, the closer neighbors are weighted far more heavily than those further away. This has the effect of effectively reducing the number of neighbors that contribute to the class membership value of the input data point. For MUP classification m was set to 2 as Keller [53] concluded that almost equal error rates have been obtained over a wide range of values of m.

The fuzzy k-NN classifier relies on the estimation of the membership functions for the labelled reference patterns. Methods for automatic estimation of membership functions have been summarized in [70]. We are interested in the fuzzy nearest neighbour labelling techniques beside the crisp labelling. In crisp labelling, each labelled reference pattern is assigned complete membership in its class and zero membership in all other classes, i.e., $l(v_i) \in [0,1]^M$. The fuzzy nearest neighbour labelling, known as soft labelling, assigns memberships to labelled reference patterns according to the k-nearest neighbors rule. It is required to estimate M degrees of membership $(\mu_{\omega_1}(v_i), \mu_{\omega_2}(v_i), ...\mu_{\omega_M}(v_i))$ for any $v_i \in V$ by first finding the k patterns in V closest to each labelled reference pattern v_i and then calculating the membership functions.

We implemented two schemes of soft labelling. The first scheme proposed by Jóźwik [49] assigns memberships to the labelled reference patterns according to the following formula:

$$\mu_{\omega_i}(v_j) = \frac{k_i}{k} \tag{3.20}$$

The second scheme proposed by Keller et al. [53] assigns memberships to the labelled

reference patterns according to the following formula:

$$\mu_{\omega_i}(v_j) = \begin{cases} 0.51 + (\frac{k_i}{k} * 0.49), & if \quad j = i \\ (\frac{k_i}{k} * 0.49), & if \quad j \neq i. \end{cases} \tag{3.21}$$

where k_i is the number of labelled reference patterns amongst the k closest labelled reference patterns which are labelled in class ω_i, and j ranges from 1 to n. Notice that the classes share the membership, i.e.,

$$\sum_{i=1}^{M} \mu_{\omega_i}(v_j) = 1, \quad \forall v_j \in V. \tag{3.22}$$

3.5.3 Fuzzy k-NN Classifier for MUP Classification

The fuzzy k-NN classifier for MUP classification estimates a measure of assertion expressing confidence in the decision of classifying a MUP to a particular train. It determines for each candidate MUP j a class membership $\mu_i(j)$ calculated from (3.18) representing the shape-based strength of membership of MUP j in MUPT i and a firing assertion decision function $A^j_{FA_i}$ assessing the time of occurrence of MUP j with respect to the established firing pattern of MUPT i.

The firing pattern information is represented by the firing assertion decision function A_{FA} similar to as is done for the Certainty classifier. For candidate MUP j and MUPT i, $A^j_{FA_i}$ is evaluated by:

$$A^j_{FA_i} = A_f(I^j_{bi}, \mu_i, \sigma_i) \ . \ A_f(I^j_{fi}, \mu_i, \sigma_i) \tag{3.23}$$

where, $A_f(I, \mu, \sigma)$ is a firing time assertion function based on the deviation of an IDI, I, from the estimated mean IDI, μ, of a train that has an estimated standard deviation, σ and I_{bi} and I_{fi} are forward and backward IDIs, respectively, all as defined above for the Certainty classifier (see equations (3.9), (3.10), and (3.11)).

The overall assertion value for assigning MUP j to MUPT i (A^j_i) is defined as:

$$A^j_i = \mu_i(j) \ . \ A^j_{FA_i} \tag{3.24}$$

MUP j is assigned to the train with the highest assertion value and if this value is above the minimum assertion value threshold A_m of the train to which a classification is to be made, otherwise MUP j is left unassigned.

The value of class membership $\mu_i(j)$ and the firing time assertion decision function $A^j_{FA_i}$ is restricted to the interval [0, 1] and it corresponds to the confidence in the classification given the information of each function. A value of 1 corresponds to the ideal situation (maximum assertion) with respect to the information relevant to that function. However, for a classification that is certainly incorrect with respect to any source of information the corresponding function yields a value approaching 0.

3.5.4 Adaptive Fuzzy k-NN Classifier

The adaptive fuzzy k-NN classifier (AFNNC) uses an adaptive assertion-based approach for assigning MUPs to trains. The similarity criterion for grouping MUPs is based on a combination of MUP shapes and an active and passive use of MU firing patterns.

The adaptive nature of MUP classification is related to the adjustment of the minimal assertion threshold for each MUPT based on train firing pattern statistics and it follows the algorithm described in Section 3.3.2.

The classification pass during the first stage assigns each candidate MUP j to the train that has the highest shape assertion of belonging to it expressed in terms of the maximum membership value calculated from (3.18) such that it is assigned to the train given by:

$$\omega(j) = arg \max_{i=1}^{M}(\mu_{\omega_i}(j)) \tag{3.25}$$

if it is higher than the shape assertion threshold initially set to $A_m = 0.6$.

The classification passes in the second stage assign each candidate MUP j to the train that has the best match of MUP j shape and expected firing time and if this match has an assertion above the current assertion threshold for the train. The overall assertion of a MUP belonging to a train is given by (3.24) and measured in terms of the multiplicative

combination of the membership function value calculated from (3.18) and the firing time assertion (3.23). If the overall assertion is higher than the minimum assertion threshold initially set to $A_m = 0.02$, the candidate MUP j is assigned to the train given by:

$$\omega(j) = arg \max_{i=1}^{M}(A_i^j) \tag{3.26}$$

The adjustment of the minimal assertion threshold follows the adaptive setting of MUPT assignment threshold algorithm described in Section 3.3.2.

3.6 Matched Template Filtering Classification

The basic MUP matched template filtering algorithm consists of sliding MUPT templates over the EMG signal detected MUPs, and for each candidate MUP calculating a distortion, or correlation, measure estimating the degree of dissimilarity, or similarity, between the template and the MUP. Then, the minimum distortion, or maximum correlation, position is taken to represent the instance of the template into the signal under consideration, with a threshold on the similarity/dissimilarity measure allowing for rejection of poorly matched MUPs. In this thesis we used correlation measure as estimates of the degree of similarity between a MUP and MUPT templates.

The correlation between two signals represents the degree to which signals are related, and cross correlation analysis enables the degree of waveform similarity between two different signals to be determined. It provides a quantitative measure of the relatedness of two signals as they are progressively shifted in time with respect to each other.

In this section, we present two matched template filters for supervised MUP classification during EMG signal decomposition. They are: the normalized cross correlation which is the most widely used correlation measure [102] and a pseudo-correlation [32] measure.

3.6.1 Matched Template Filter Classifier for MUP Classification

The matched template filter (MTF) classifier for MUP classification estimates a measure of similarity between a candidate MUP j and the MUPT templates expressing confidence in the decision of classifying a MUP to a particular train. It determines for each candidate MUP j a normalized cross correlation value calculated from (3.27) or a pseudo correlation value calculated from (3.28) representing the strength of resemblance of the MUP j with the MUPT templates:

$$NCC^{j}_{\omega_i}(x) = \frac{\sum_{k=1}^{n} m(x+k).\ T_i(k)}{\sqrt{\sum_{k=1}^{n} m(x+k)^2}.\ \sqrt{\sum_{k=1}^{n} T_i(k)^2}} \tag{3.27}$$

$$pC^{j}_{\omega_i}(x) = \frac{\sum_{k=1}^{n} \left(T_i(k).\ m(x+k) - |T_i(k) - m(x+k)|.\ \max\{|T_i(k)|,\ |m(x+k)|\}\right)}{\sum_{k=1}^{n} (\max\{|T_i(k)|,\ |m(x+k)|\})^2} \tag{3.28}$$

Denote ρ to be the matched template filter correlation such that:

$$\rho_{\omega_i}(x) = \begin{cases} NCC^{j}_{\omega_i}(x), & \textit{when choosing normalized cross correlation,} \\ pC^{j}_{\omega_i}(x), & \textit{when choosing pseudo correlation.} \end{cases} \tag{3.29}$$

where, $x = 1, 2, ..., n$ is the shifting position and n is the dimension of the feature vector, m is the candidate MUP j feature vector, and T_i is the MUPT i template feature vector. Figure 3.8 shows the similarity between a candidate MUP, drawn in a solid line, and a MUPT template, drawn in a dashed line, along with the degree of similarity in terms of the normalized cross correlation and the pseudo correlation measures.

The MTF classifier also determines for MUP j a firing time similarity decision function $S^{j}_{FS_i}$ with respect to the established firing pattern of the train.

The firing pattern information is represented by the firing similarity decision function S_{FS} similar to as is done for the Certainty classifier. For candidate MUP j, $S^{j}_{FS_i}$ is evaluated by:

$$S^{j}_{FS_i} = S_f(I^{j}_{bi}, \mu_i, \sigma_i)\ .\ S_f(I^{j}_{fi}, \mu_i, \sigma_i) \tag{3.30}$$

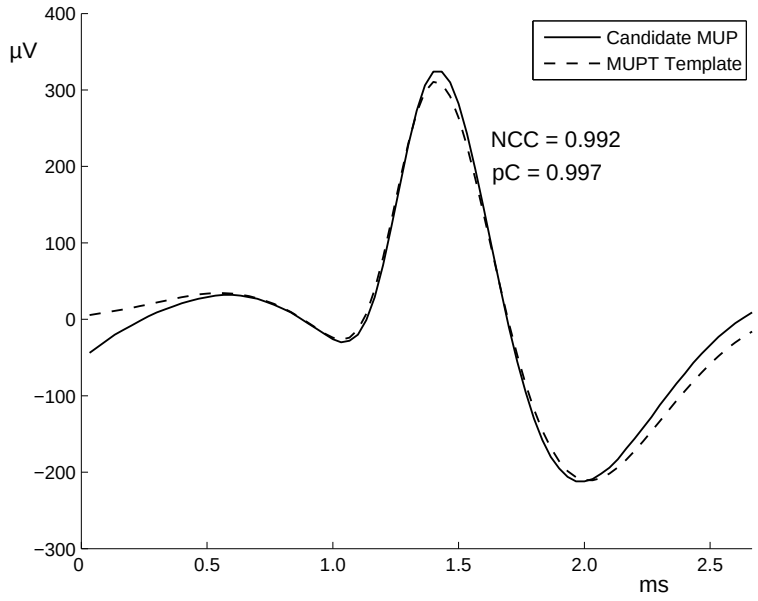

Figure 3.8: Illustration of similarity between a candidate MUP and a MUPT template.

where, $S_f(I, \mu, \sigma)$ is a firing time function based on the deviation of an IDI, I, from the estimated mean IDI, μ, of a train that has an estimated standard deviation, σ and I_{bi} and I_{fi} are forward and backward IDIs respectively, all as defined above for the Certainty classifier (see equations (3.9), (3.10), and (3.11)).

The overall similarity measure for assigning MUP j to MUPT i (S_i^j) is given by the multiplicative combination of $\rho_i(j)$ and $S_{FS_i}^j$ defined in (3.31) as:

$$S_i^j = \rho_i(j) \; . \; S_{FS_i}^j \tag{3.31}$$

MUP j is assigned to a train with the highest similarity measure and if it is above the minimal similarity threshold S_m of the train to which a classification is to be made, otherwise MUP j is left unassigned.

The value of similarity measure $\rho_i(j)$ and the firing time similarity decision function $S_{FS_i}^j$ is restricted to the interval [0, 1] and it corresponds to the confidence in the classification given the information of each function. A value of 1 corresponds to the ideal situation (maximum similarity) with respect to the information relevant to that function. However,

for a classification that is certainly incorrect with respect to any source of information the corresponding function yields a value approaching 0.

3.6.2 Adaptive Matched Template Filter Classifier

The adaptive matched template filter classifier (AMTF) uses an adaptive similarity approach for assigning MUPs to trains. The similarity criterion for grouping MUPs is based on a combination of MUP shapes and an active and passive use of MU firing patterns.

The adaptive nature of MUP classification is related to the adjustment of the minimal similarity threshold for each MUPT based on train firing pattern statistics and it follows the algorithm described in Section 3.3.2.

The classification pass during the first stage assigns each candidate MUP j to the train that has the highest shape similarity measure of belonging to it expressed in terms of the maximum correlation value calculated from (3.27) or (3.28) such that it is assigned to the train given by:

$$\omega(j) = arg \max_{i=1}^{M}(\rho_{\omega_i}(j)) \tag{3.32}$$

if it is higher than the shape similarity threshold initially set to $S_m = 0.6$.

The classification passes in the second stage assign each candidate MUP j to the train that has the best match of MUP j shape and expected firing time if this match has a similarity measure above the current similarity threshold for the train. The similarity measure of a MUP belonging to a train is given in terms of the multiplicative combination of the correlation value calculated and the firing time similarity. If the similarity measure is higher than the similarity threshold initially set to $S_m = 0.02$, the candidate MUP j is assigned to the train given by:

$$\omega(j) = arg \max_{i=1}^{M}(S_i^j) \tag{3.33}$$

The adjustment of the minimal similarity threshold follows the adaptive setting of MUPT assignment threshold algorithm described in Section 3.3.2.

Chapter 4

EMG Signal Decomposition Classifier Fusion Model

4.1 Introduction

Multi-classifier decision level fusion can be defined as the process of fusing information from individual classifiers after each classifier has given an opinion about the recognition of a pattern. The output information from the multiple classifiers may agree or conflict with each other and the task of the designed classifier fusion system becomes the search for classifiers, from the pool of base classifiers, having the maximum degree of agreement.

The EMG signal decomposition process can be considered a classification problem that abounds uncertainty. Our goal in this thesis is the management of the uncertainty in classifying MUP patterns through the fusion of the decisions of multiple classifiers. Instead of designing a high performance classifier, we designed and constructed a number of classifiers using different methodologies and different features that can complement each other. Each classifier itself need not have an excellent performance, but the appropriate choice and combination of these individual classifiers should produce a highly reliable performance.

This chapter describes the developed classifier fusion model architecture and the stages

employed for designing a classifier fusion system: the design of the classifier ensemble and the design of the aggregation module; along with the classifier fusion schemes used.

4.2 Classifier Fusion Model Architecture

The developed system deals with EMG signal decomposition and considers it as a classification problem by implementing a multiple classifier fusion system.

The system architecture implemented in this thesis belongs to the parallel category, where all the individual base classifiers are invoked concurrently and independently, and an aggregator combines their output results as shown in Figure 4.1.

This section describes the system architecture and specifies its base classifiers, model input space, aggregation module, and data transformation when the model includes different types of classifiers.

4.2.1 Multiple Classifier Fusion System

The multiple classifier system shown in Figure 4.1 consists of a set of individual classifiers and an aggregator, which combines the results of the individual classifiers to produce the final classification decision. Various combination configurations can be obtained by using different classifier architectures, different input features, or different aggregator characteristics.

The thesis model employs a heterogeneous combination approach by combining classifier ensembles of different kinds. The first ensemble represents a set of certainty-based classifiers consisting of the Certainty classifiers [75], [99] and/or adaptive certainty classifiers [85]. The second ensemble represents a set of assertion-based classifiers consisting of the fuzzy k-NN and/or adaptive fuzzy k-NN classifiers [83]. Whereas the third ensemble represents a set of matched template filter classifiers consisting of the matched template filter and/or adaptive matched template filter classifiers.

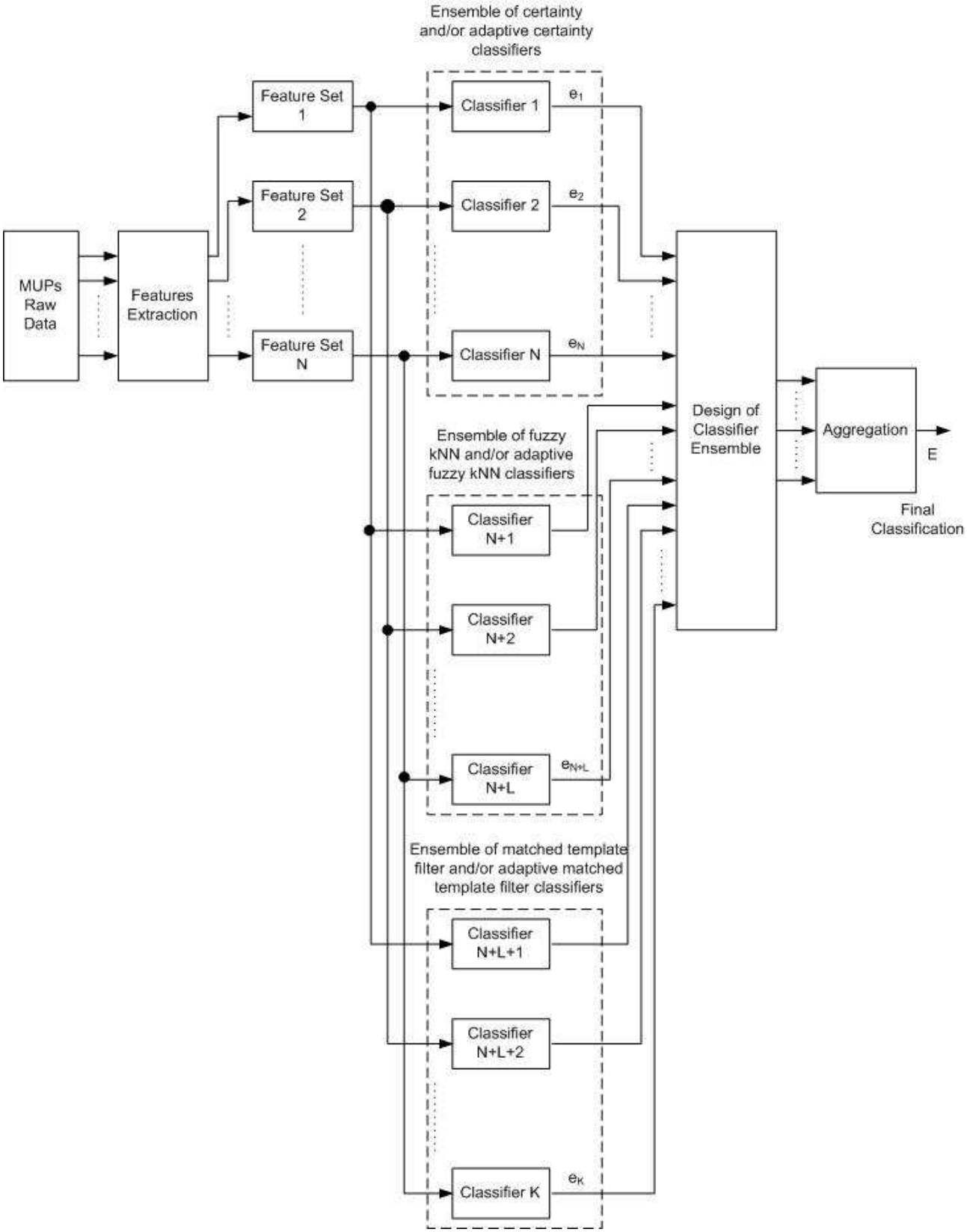

Figure 4.1: Developed classifier fusion model architecture.

4.2.2 Model Base Classifiers

The model base classifiers, in an abstract form, may be considered as systems that process a MUP feature vector and provide both a decision representing the train (class) label that the MUP belongs to and a numeric value representing the overall confidence in the MUP assignment. If the overall confidence associated with the classification of a MUP exceeds an assignment threshold then the MUP assignment to the train will be accepted. Otherwise, if the confidence is below the assignment threshold, the MUP pattern is left unassigned.

Generally, the assignment rate of the classifier depends on the internal parameters of the classifier. In the base classifiers, by varying the minimum assignment threshold one obtains an accuracy as a function of the assignment rate. Low values of the assignment threshold correspond to classifications that are not highly certain, i.e., with higher assignment rates and higher error rates. Whereas higher values of the assignment threshold correspond to classifications which are quite certain with more conservative assignment and hence lower assignment rates but with lower error rates.

To improve the overall classification performance, we may make decisions, concerning the assignment of a candidate MUP, based on the opinion of different classifiers, in analogy with decisions taken for real world problem solutions, where usually decisions made by teams are better than decisions made by individuals.

4.2.3 Model Input Space

To build a pool of base classifiers, we consider, as stated in Sections 3.1 and 4.2.1, ensembles composed of different kinds of classifiers and fed with different input features.

The different kinds of base classifiers are stated in Sections 3.1 and 4.2.1. The model follows the so-called *overproduce and choose* [36], [76] strategy and the *test and select* [92] approach to ensemble combination by allowing the generation of a large set of candidate classifiers and then the choice, i.e., the candidate classifiers to be combined, is based on the selection of classifiers used in the combination.

For each kind of base classifier, there are approaches to get different classifier configurations:

1. Using different seeding data: The model uses three methods to specify the seeding MUP data: clustering data, sequential assignment data, and high certainty data. Details for extracting each type of seeding data are described in Section 5.2.6.

2. Changing the order of input vectors presentation to the classifiers. Where we may use either an ordering based on a MUP's time of occurrence within the EMG signal or may use a random order due to the observation that the base classifiers are sensitive to the sequence of input vectors and yield different knowledge in the form of different assignments and accuracy for different orderings of the input set.

3. Varying classifier parameter values such as the minimum confidence threshold.

A second possible way of building a set of base classifiers is through using classifiers with different input features extracted from the raw data. For this purpose, we may build classifiers that receive raw data, filtered data using low-pass differentiators of first-order and second-order, and wavelet-domain features based on the discrete wavelet transform of the MUP waveforms with different types of wavelets and using different scales. Details of feature extraction are presented in Section 5.2.1.

4.2.4 Aggregation of Multiple Decisions

In order to combine the decisions of multiple classifiers, it is necessary that all classifiers provide outputs on the same information level. The developed model may combine the ensembles of classifiers based on two levels: the measurement level and the abstract level. The measurement level has the most information and the abstract level has the least information.

To combine the outputs from the multiple disparate classifiers based on the abstract level, the voting schemes are used, where the target class that receives the highest number

of votes is selected as the final predicted class.

If the combination is based on the measurement level, methods that do not require prior training are used such as combination schemes based on the product, sum, max, min, median, and average rules.

Beside using the previous fixed combination rules we also used trainable combination methods. A training set can be used to adapt the combining classifier to the classification problem, where in this case the combination operator also functions as a classifier. The outputs of the base classifiers can be used as the input features of a general classifier used for combining. The fuzzy integral method is one of the trainable combiners that has been studied in this thesis.

4.2.5 Base Classifier Output Data Transformation

For the abstract level of combination, the confidence output generated when classifying any candidate MUP by any base classifier does not affect the combined decision. Whereas for the measurement level of combination, the confidence output plays a critical role in classification and erroneous information coming from any base classifier can be much more easily compensated for with the correct output of another base classifier.

As stated in Section 4.2.1, the thesis model employs a heterogeneous combination approach by combining different kinds of classifier ensembles. The model measurement level combination creates a major problem due to the incomparability of base classifier outputs. Certainty classifiers and adaptive certainty classifiers provide outputs in terms of total certainty based on (3.12), fuzzy k-NN and adaptive fuzzy k-NN classifiers provide outputs in terms of total assertion based on (3.24), and matched template filter and adaptive matched template filter provide outputs in terms of total similarity based on (3.31). All classifiers provide outputs only for the closest and the second closest MUPTs and zero for the other MUPTs. Therefore, the scales of the outputs from the different base classifiers are incomparable and need preprocessing before combination.

To deal with the different meanings and scales of the base classifier outputs, a preprocessing operation in the form of a data transformation becomes essential before combining the classifier outputs. The preprocessing operation used is based on normalization of the classifier outputs so that they satisfy the axioms of probability [2]. The following two transformation functions were studied:

1. Output sum normalization function:

$$T_1 : e_{ik} \rightarrow t_{ik} \tag{4.1}$$

$$where \;\; t_{ik} = \frac{e_{ik}}{\sum_{i=1}^{M} e_{ik}} \tag{4.2}$$

2. Output square sum normalization function:

$$T_2 : e_{ik} \rightarrow t_{ik} \tag{4.3}$$

$$where \;\; t_{ik} = \frac{(e_{ik})^2}{\sum_{i=1}^{M} (e_{ik})^2} \tag{4.4}$$

where, $i = 1, 2, ..., M$; and $k = 1, 2, ..., K$ for M classes and K classifiers.

After this normalization process is applied, all the outputs of the classifiers are transformed into the normalized values as follows:

$$\sum_{i=1}^{M} t_{ik} = 1, \quad 0 \leq t_{ik} \leq 1.0, \; \forall \; i, \; \forall \; k. \tag{4.5}$$

4.3 Design of Classifier Fusion System

The overall accuracy of the combined classifier depends not only on the way the base classifiers are fused but also on the selection of the classifiers used in the fusion. Accordingly, the design of classifier fusion systems involves two main stages:

1. The design of the classifier ensemble,

2. The design of the aggregation module.

To design the most appropriate classifier fusion system for MUP classification, we follow the so-called *overproduce and choose* [36], [76] paradigm (also called the *test and select* [92] approach). The basic idea is to produce an initial large set of candidate classifier ensembles, and then select the ensemble that can be combined to achieve better accuracy.

Figure 4.2 shows the main phases involved in the classifier fusion system design cycle: the base classifier overproduction phase, the ensemble choice phase, the aggregator design phase, and the performance evaluation phase. The classifier ensembles or aggregation module must be redesigned when the output of the performance evaluation phase is not satisfactory, and in accordance to that, Figure 4.2 shows feedback from later design phases to the earlier ones.

The overproduction design phase produces a large set of base classifiers and in Section 4.2.3 we stated how we can produce different candidate classifiers.

The ensemble choice phase selects the subsets of classifiers that can be combined to achieve better accuracy. The subset giving the best accuracy could be obtained by exhaustive enumeration. Such that, if K is the size of the base classifier set produced by the overproduction phase, the number of possible subsets is equal to $\sum_{i=1}^{K} \binom{K}{i}$. Therefore, there is a need for a strategy to limit the computational complexity of the choice phase and follow techniques that choose an effective classifier ensemble without hypothesizing a specific combination rule [86]. Accordingly, techniques to evaluate the error diversity of classifiers that make up an ensemble have been used for classifier selection purposes. For this purpose, we chose the kappa statistic to select base classifiers having an excellent level of agreement to form ensembles giving satisfactory classification performance.

In the aggregator design phase, the choice of the combination function should take into account the dependency among classifiers. In actual practice, a trial and error procedure is performed, because a clear model of the dependency among classifiers is difficult to obtain [86]. If all classifiers in an ensemble are totally positively dependent, i.e., they are

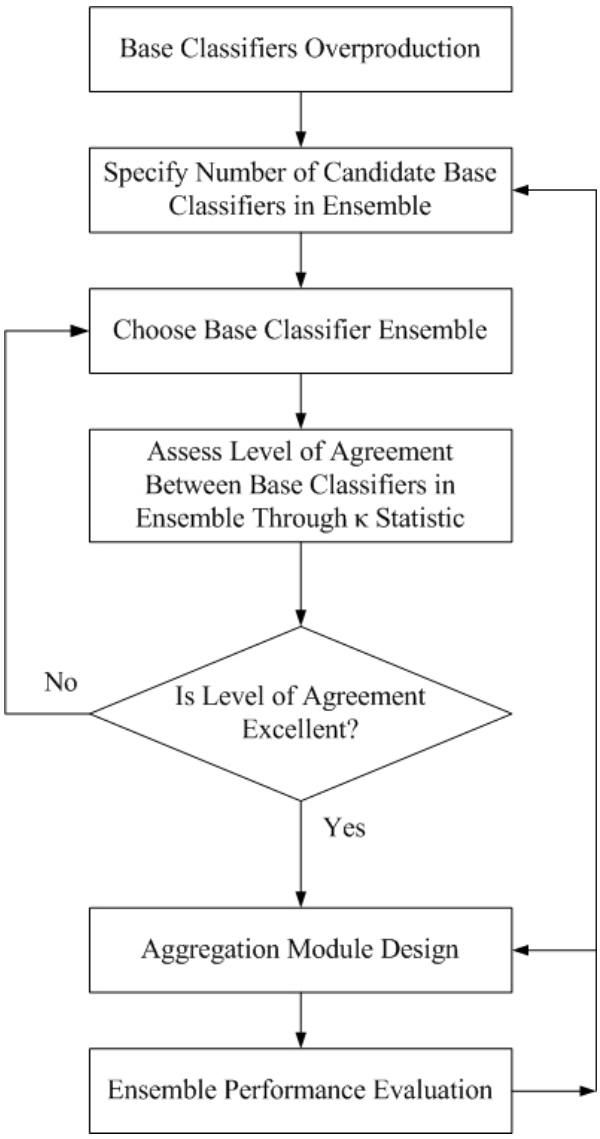

Figure 4.2: Classifier fusion system design cycle.

identical, there be will no improvement in performance. However, if there are negatively dependent, i.e., the base classifiers commit mistakes on different MUP patterns, we could expect improvement in performance. This raises the issue of quantifying the degree of agreement among dependent classifiers.

Performance evaluation is performed by assessing the classification accuracy of the selected classifier ensemble using the classifier fusion module designed in the previous phase.

4.3.1 Independence of Classifiers

When combining identical classifiers, we do not gain any improvement in performance. Therefore, the key issue in building multiple classifier systems and the system property responsible for team strength is the diversity among the team of classifiers.

Diversity in combining classifiers refers to the following terms [61], [93]: independence, orthogonality, and complementarity.

Orthogonality is used to denote the tendency of classifiers to make different decisions. Since classifiers may have different strengths and weaknesses, combining them is assumed to have a compensatory or complementary effect.

A classifier combination is especially useful if the individual classifiers are largely independent where independence between individual classifiers is typically viewed as an asset in classifier fusion [59]. This may be guaranteed by the use of classifiers of different kinds, and based on different training sets. Also, different feature sets may be used to explicitly force the individual classifiers to use independent information.

In general, classifiers are considered independent if they produce independent errors but in classifier fusion systems, despite the fact that they may work independently, their errors are usually strongly dependent since they act on the same data.

4.4 The Design of the Classifier Ensemble

The improvement in performance of a classifier fusion system relies on properly choosing the base classifiers to be fused. As stated in Section 4.3, choosing base classifiers can be performed directly through exhaustive search with the performance of the fusion being the objective function. As the number of base classifiers increases, this approach becomes computationally too expensive.

Instead of the exhaustive search method of choosing base classifiers, in this section we are interested in exploiting the diversity measure for designing classifier teams. Diversity, as a measure, has been used for selecting ensembles in design of multiple classifier systems [87]. Shipp and Kuncheva [93] summarize 10 measures of diversity for pairwise and non-pairwise base classifiers.

To evaluate the diversity, an appropriate measure is needed [58]. In this thesis, we chose the kappa measure κ to estimate the level of agreement between the base classifier outputs [31], i.e., to measure the degree of decision similarity between the base classifier outputs, for the following reasons:

> κ depends on the individual accuracies of the classifiers, and has a specific value 0 for statistically independent classifiers. κ varies between -1 and +1. κ close to 1 indicates that the classifiers agree in the recognition of the same MUP patterns and κ = -1 means that the assignment by the classifiers are different for the same MUP patterns.

4.4.1 Assessing Base Classifiers Agreement

First we consider the case of two classifiers such that each MUP pattern of an EMG signal consisting of N MUPs is classified independently by the two classifiers into M MUPTs. It

Table 4.1: 2 x 2 Table of dichotomous outcome for two classifiers.

MUPs correctly classified by both classifiers - N_{cc}	MUPs correctly classified by the first and incorrectly by the second - N_{cw}
MUPs incorrectly classified by the first and correctly by the second - N_{wc}	MUPs incorrectly classified by both classifiers - N_{ww}

is desired to measure the degree of agreement on each MUPT separately as well as across all MUPTs.

Initially, we will focus on assessing agreement on correct classification and agreement on error rather than agreement in general. In order to do so we match the results from individual classifiers for a given EMG signal in a dichotomous outcome based on the correct or erroneous outcome of the classification by two classifiers. This results in a 2 x 2 table shown in Table 4.1, where c refers to correct and w refers to wrong.

Let e_i and e_j be the two base classifiers, and consider that when e_i and e_j propose the true MUPT for a candidate MUP, they agree. If they both propose incorrect MUPTs, they also agree. In all the other situations, they disagree. Putting the entries of Table 4.1 in proportions of the total number of MUPs N, we get the diversity matrix representing the percentage of agreement and disagreement between the two classifiers e_i and e_j shown in Table 4.2.

The pairwise agreement between classifiers e_i and e_j can be obtained by the kappa statistic given by [31]:

$$\kappa = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} \tag{4.6}$$

The kappa statistic was first proposed by Cohen [16] and it has the same interpretation as the *intraclass correlation coefficient* [31]. κ expresses a special type of relationship between classifiers and it quantifies the level to which the classifiers agree in their decisions beyond

Table 4.2: Diversity matrix DM_{ij} of e_i and e_j classifiers.

	e_j correct	e_j wrong	Total
e_i correct	a	b	$p_1 = a + b$
e_i wrong	c	d	$q_1 = c + d$
Total	$p_2 = a + c$	$q_2 = b + d$	$a + b + c + d = 1$

a, b, c, d are proportions, such that $a = \frac{N_{cc}}{N}$, $b = \frac{N_{cw}}{N}$, $c = \frac{N_{wc}}{N}$, $d = \frac{N_{ww}}{N}$, and p_i, q_i are marginal probabilities.

any agreement that could occur due to chance.

The kappa statistic corrects for chance expected agreement meaning that it should reflect the amount of agreement in excess of what would be expected by chance. It not only gives a measure of the degree of agreement, but it also has a test associated with it that can be employed to check if the apparent agreement cannot be attributed to chance only. It is also helpful that the kappa statistic can show the level of agreement. A value of 0.40 is considered to represent poor agreement beyond chance, values between 0.40 and 0.75 indicate fair agreement, and values beyond 0.75 indicate excellent agreement [31].

Agreement can also be assessed on a per MUPT basis for two base classifiers. If M MUPTs are available, then we will have an $(M+1)$ x $(M+1)$ diversity matrix noting that $M+1$ represents the unassigned category. However, this matrix indicates how a pair of classifiers classify in relation to one another and not the reference data. In this case what is taken into account is not whether the classification is correct or not, but only if the base classifiers agree or not in their classification of a particular MUP. Then the kappa statistic will be computed using the kappa hat statistic formula for multiple outcomes given by

(4.7) [6]:

$$\hat{\kappa} = \frac{N \sum_{i=1}^{M+1} d_{ii} - \sum_{i=1}^{M+1} d_{i+} \ . \ d_{+i}}{N^2 - \sum_{i=1}^{M+1} d_{i+} \ . \ d_{+i}} \tag{4.7}$$

where d_{ii} is the number of MUPs in row i and column i of the diversity matrix, d_{i+} and d_{+i} are the marginal totals for row i and column i, respectively, and N is the total number of MUPs.

The above analysis can be extended to the case of more than two classifiers. For an ensemble of K base classifiers $e_k, k = 1, 2, ..., K$, known to be dependent on each other as they work on the same data, used to classify a set of N MUP patterns into M MUPTs and the unassigned category $\omega_i \in \Omega = \{\omega_1, \omega_2, ..., \omega_M, \omega_{M+1}\}$, we want to estimate the strength of the association among them through measuring the degree of agreement among dependent classifiers. For $j = 1, 2, ..., N; i = 1, 2, ..., M + 1$ denote by d_{ji} the number of classifiers which assign candidate MUP m_j to class ω_i, i.e.,

$$d_{ji} = \sum_{k=1}^{K} T(e_k(m_j) = \omega_i) \tag{4.8}$$

where $T(e = \sigma)$ is a binary characteristic function and it equals 1 if $e = \sigma$ and 0 otherwise. Note that $\sum_{i=1}^{M+1} d_{ji} = K$ for each MUP m_j. Table 4.3 shows the per MUP pattern diversity matrix through d_{ji}.

Based on the per MUP pattern diversity matrix of K classifiers, the degree of agreement among dependent classifiers $e_k(x), k = 1, 2, ..., K$ is measured using the following kappa hat statistic formula for multiple outcomes and multiple classifiers [31]:

$$\hat{\bar{\kappa}} = 1 - \frac{NK^2 - \sum_{j=1}^{N} \sum_{i=1}^{M+1} d_{ji}^2}{KN(K-1) \sum_{i=1}^{M+1} \bar{p}_i \ \bar{q}_i} \tag{4.9}$$

where $p_i = \frac{\sum_{j=1}^{N} d_{ji}}{NK}$ represents the overall proportions of outputs of classifiers in MUPT ω_j, and $q_i = 1 - p_i$. The value of the kappa hat statistic $\hat{\kappa}_i$ for MUPT $\omega_i, i = 1, 2, ..., M$ and the unassigned category ω_{M+1} may be measured using the formula [31]:

$$\hat{\kappa}_i = 1 - \frac{\sum_{j=1}^{N} d_{ji}(K - d_{ji})}{KN(K-1)\bar{p}_i\bar{q}_i} \tag{4.10}$$

Table 4.3: Per MUP pattern diversity matrix of K classifiers.

MUP Pattern	MUPT ω_1	MUPT ω_2	...	MUPT ω_M	MUPT ω_{M+1}	$\sum_{i=1}^{M+1} d_{ji}^2$
m_1	d_{11}	d_{12}	...	d_{1M}	$d_{1(M+1)}$	$\sum_{i=1}^{M+1} d_{1i}^2$
m_2	d_{21}	d_{22}	...	d_{2M}	$d_{2(M+1)}$	$\sum_{i=1}^{M+1} d_{2i}^2$
.	.	.	...	.	.	.
.	.	.	...	.	.	.
.	.	.	...	.	.	.
m_N	d_{N1}	d_{N2}	...	d_{NM}	$d_{N(M+1)}$	$\sum_{i=1}^{M+1} d_{Ni}^2$
Total	$\sum_{j=1}^{N} d_{j1}$	$\sum_{j=1}^{N} d_{j2}$	...	$\sum_{j=1}^{N} d_{jM}$	$\sum_{j=1}^{N} d_{j(M+1)}$	$\sum_{j=1}^{N} \sum_{i=1}^{M+1} d_{ji}^2$

In this thesis, we proposed a diversity-based hybrid classifier approach consisting of an aggregator module that contains two combiners and with each combiner there is a pre-stage classifier selection module responsible for selecting the base classifiers comprising the ensemble. The selection of base classifiers to be combined by the first combiner is based on (4.9) where the set of base classifiers having the maximum level of agreement are chosen. Whereas, the selection of base classifiers to be combined by the second combiner is based on (4.10) where the set of base classifiers having the minimum level of agreement considering only the unassigned category are chosen, i.e., when $i = M + 1$.

4.5 The Design of the Aggregation Module

As stated in Section 1.4.3, our focus in this thesis is on classifier fusion and based on that a parallel configuration as shown in Figure 1.3 and Figure 4.1 is used. Our interest in this section is the design of the aggregation module shown in the configuration.

The aggregator in a classifier fusion system combines the base classifier outputs to achieve a group consensus. Aggregators may be data independent [50], where they don't show any dependence on the data and they solely rely on the output of base classifiers to produce a final classification decision irrespective of the pattern being classified. Based on (2.6) and for a set of decision confidences $\{Cf_{ik}(x), i = 1, 2, ..., M; k = 1, 2, ..., K\}$ for M classes and K classifiers, the data independent combining scheme outcome for class ω_i takes the form:

$$Q_i(x) = F_k(Cf_{ik}(x)) \tag{4.11}$$

where $\{Q_i(x), i = 1, 2, ..., M\}$ are the combined classifier decision confidences using the mapping F_k of the combining approach.

Aggregators may be data dependent [50] with implicit or explicit dependency on data. Implicit data dependent combining schemes take the form:

$$Q_i(x) = F_k(W_{ik}(e_{ik}(x)), Cf_{ik}(x)) \tag{4.12}$$

where the weights, W_{ik}, assigned to any output depend on the outcome $e_k(x)$ of the base classifier. Whereas explicit data dependent combining schemes take the form:

$$Q_i(x) = F_k(W_{ik}(x), Cf_{ik}(x)) \tag{4.13}$$

where the weights, W_{ik}, depend on the input pattern x.

The thesis developed system combines heterogeneous sets of base classifier ensembles of different kinds and employs either a one level classifier fusion scheme or a hybrid classifier fusion approach. The hybrid aggregation module is a hybrid combination of two classifier fusion schemes: the first is based on the abstract level and the second is based on the measurement level. Both combiners may be either data independent or the first combiner data independent and the second data dependent. For the purpose of this thesis, we used as first combiner the majority voting scheme, while we used as the second combiner one of the fixed combination rules behaving as data independent combiner or used the fuzzy integral with the λ-fuzzy measure as an implicit data dependent combiner.

4.5.1 Majority Voting Classifier Fusion Scheme

To combine the outputs from the multiple disparate base classifiers based on the abstract level, the majority voting scheme was used. For majority voting, the mapping function F_k in (4.11) would be a nonlinear mapping [50].

According to the majority voting generalized expression given in (2.4), the fraction of the agreed classifiers α varies with each MUP. The target MUPT that receives the highest number of votes was selected as the final predicted train provided that over half of the classifiers agreed on that MUPT.

4.5.2 Fixed Combination Rules

The fixed combination rules [1], [23], [55] are data independent aggregators and do not require prior training and are used for combining the set of decision confidences $\{Cf_{ik}(x), i = 1, 2, ..., M; k = 1, 2, ..., K\}$, interpreted by (2.6), for M classes and K base classifiers $\{e_k(x), k = 1, 2, ..., K\}$ into a combined classifier decision confidences $\{Q_i(x), i = 1, 2, ..., M\}$. They are the *product*, the *sum*, the *max*, the *min*, the *median*, and the *average* rules. The combined decision confidence $Q_i(x)$ for class ω_i is computed by:

$$Q_i(x) = rule_k \{Cf_{ik}(x)\} \tag{4.14}$$

where *rule* in (4.14) represents one of the fixed combination rules. Normalization may be required and formula (4.14) becomes:

$$\grave{Q}_i(x) = \frac{Q_i(x)}{\sum_i Q_i(x)} \tag{4.15}$$

The final classification is made by:

$$\omega(x) = arg \ \max_{i=1}^{M}(\grave{Q}_i(x)) \tag{4.16}$$

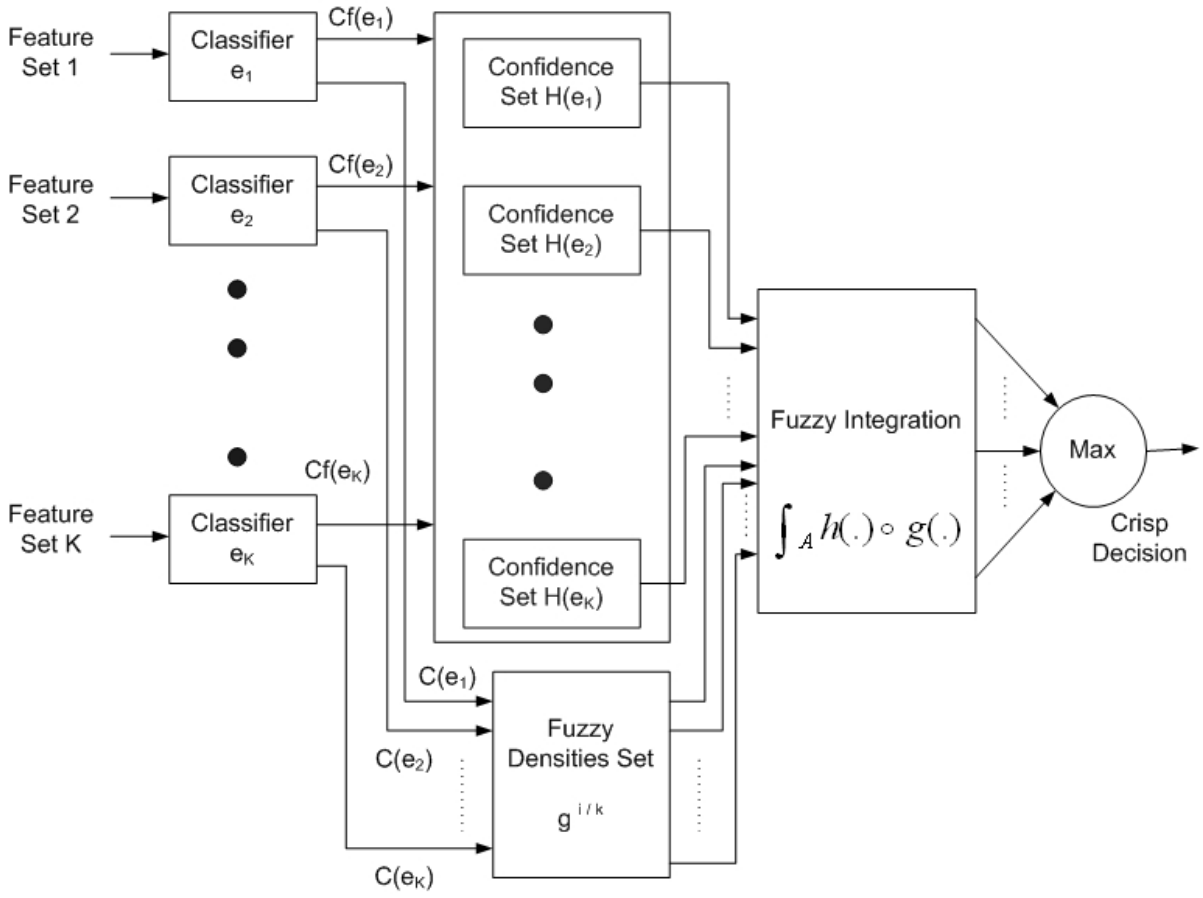

Figure 4.3: Combination of base classifiers using a fuzzy integral.

4.5.3 Fuzzy Integral Classifier Fusion Scheme

Based on the measurement level of classifier fusion, and instead of using fixed combination rules, one can train an arbitrary classifier using the M x K decision confidences $Cf_{ik}(x)$ (for all i and all k) as features in the intermediate space [23], [24].

The fuzzy integral [104], [105] is a nonlinear functional defined with respect to a fuzzy measure, a generalization of a probability measure, specifically a g_λ-fuzzy measure. It was used as a numeric-based aggregation connective approach for combining multiple classifiers [11], [12], [13] to reach a collective decision as illustrated in Figure 4.3. In this thesis, we implemented two fuzzy integral methods: the Sugeno [105] fuzzy integral and the Choquet [34] fuzzy integral.

Fuzzy integral methods utilize information concerning the worth or confidence in subsets of information sources in the decision making process [52] represented by a fuzzy measure. In the classifier fusion process, fuzzy integrals combine objective evidence, supplied by the base classifiers in the form of confidence measures, for a hypothesis with a prior expectation of the worth (fuzzy density values) of subsets of these classifiers.

Let $\Omega = \{\omega_1, \omega_2, ..., \omega_M\}$ be the set of MUPTs into which MUP patterns will be classified. Let $E = \{e_1, e_2, ..., e_K\}$ be the set of base classifiers, and x be the MUP pattern under consideration for classification. Let $H(e_k) : E \rightarrow [0, 1]$ be the confidence set of base classifiers e_k containing the partial evaluation of the MUP pattern x for classes set Ω, i.e., $H(e_k) = \{h_1(e_k), h_2(e_k), ..., h_M(e_k)\}$, such that $h_i(e_k)$ is an indication of the confidence in MUP pattern x classification to class ω_i using classifier e_k, where 1 indicates absolute certainty that MUP pattern x belongs to class ω_i and 0 implies absolute certainty that it does not belong to class ω_i.

Corresponding to each base classifier e_k, the degree of confidence, $g^{i/k}$, of how accurate base classifier e_k is in the recognition of class ω_i must be given. The degree of confidences $g^{i/k}$ are called fuzzy densities and can be subjectively assigned by an expert, or determined via some statistical measurements on a training set. However, these methods require some sort of prior knowledge about the information sources or require assumptions such as a Gaussian distribution of the training data. Other approaches have involved exhaustive or heuristic search methods to find density values and optimization methods to learn the entire measure [54], [106].

In this thesis, the $g^{i/k}$ densities were estimated from the training data by making them proportional to the correct classification rates of each class with each base classifier. Consider the confusion matrix of base classifier e_k denoted as $C(e_k)$, which contains the results of correctly classified and misclassified patterns. It was constructed for each base classifier and expressed in the form:

$$C(e_k) = [c_{ij}^k] \tag{4.17}$$

where $i = 1, 2, ..., M$, $j = 1, 2, ..., M + 1$, and M is the number of classes. For $i = j$, c_{ij}^k is the number of correctly classified patterns in class ω_i by classifier e_k. Whereas for $i \neq j$, c_{ij}^k is the number of patterns in class ω_i being misclassified as class ω_j by classifier e_k. Based on that, the fuzzy density values were defined as:

$$g^{i/k} = \frac{c_{ii}^k}{\sum_{j=1}^{M} c_{ij}^k} \tag{4.18}$$

Once the $g^{i/k}$'s were evaluated, the λ-fuzzy measures, $g_\lambda(A_k)$, where $A_k = \{e_1, e_2, ..., e_k\}$, were constructed for each class recursively from:

$$g_\lambda(A_1) = g_\lambda(\{e_1\}) = g^{i/1}, \quad for\ 1 \leq i \leq M \tag{4.19}$$

$$g_\lambda(A_k) = g^{i/k} + g_\lambda(A_{k-1}) + \lambda_i g^{i/k} g_\lambda(A_{k-1}), \quad for\ 1 < k \leq K\ and\ 1 \leq i \leq M \tag{4.20}$$

λ_i is calculated using formula (4.21):

$$\lambda_i + 1 = \prod_{k=1}^{K} (1 + \lambda_i g^{i/k}), \quad for\ 1 \leq i \leq M \tag{4.21}$$

This was calculated by solving the $(K - 1)$ degree polynomial and finding the unique root greater than -1.

The overall confidence for the class is the fuzzy integral value calculated using the Sugeno fuzzy integral with respect to the fuzzy measure g_λ over E [105]:

$$S = \int_A h(e) \circ g(.) = \max_{k=1}^{K} [\min(h(e_k), g(A_k))] \tag{4.22}$$

or using the Choquet fuzzy integral [34]:

$$C_g(h) = \sum_{i=k}^{K} h(e_k)\ [g(A_k) - g(A_{k+1})] \quad taking\ \ g(A_{K+1}) = 0. \tag{4.23}$$

Finally, the class ω_i with the largest fuzzy integral value is chosen as the final decision.

4.5.4 Hybrid Classifier Fusion Scheme

We propose here a hybrid type of classifier fusion scheme. It was investigated for MUP classification and across the EMG signal data sets used it had better classification performance

than applying any of the abstract or the measurement level fusion schemes individually, especially in terms of reducing the classification errors.

The hybrid classifier fusion scheme is a two-stage process and consists of two combiners: the first from the abstract level of classifier fusion represented by the majority voting data independent aggregator in the first stage and the other from the measurement level of classifier fusion represented by one of the fixed combination rules or the fuzzy integral in the second stage. It uses both combiners in a complementary manner. Figure 4.4 shows the overall basis for the hybrid classifier fusion scheme.

The hybrid fusion scheme works as follows:

1. First stage: The outputs of the ensemble candidate classifiers are presented to the majority voting combiner. If all the classifiers state a decision that a MUP is left unassigned, then there is no chance to re-assign that MUP to a valid MUPT and it stays unassigned. If over half of the classifiers assign a MUP to the same MUPT, then that MUP is allocated to that MUPT and no further assignment is processed. For these MUPs, an overall confidence value is calculated for each MUPT by averaging the confidence values given by the ensemble classifiers who contributed in the decision of assigning the MUP. In all other situations, i.e., when half or less than half of the classifiers specify a decision for a MUP to be assigned to the same MUPT, the measurement level combination scheme is used in the second stage to specify to which MUPT the MUP should be assigned based on which MUPT has the largest combined confidence value. From the first stage, a set of incomplete MUPTs are generated missing those MUPs that need to be assigned to a valid MUPT in the second stage.

2. Second stage: This stage is used for those MUPs for which only half or less than half of the ensemble classifiers in the first stage specify a decision for a MUP to be assigned to the same MUPT. The outputs of the ensemble classifiers are presented to one of the fixed rule combiners, or one of the trainable combiners represented by Sugeno

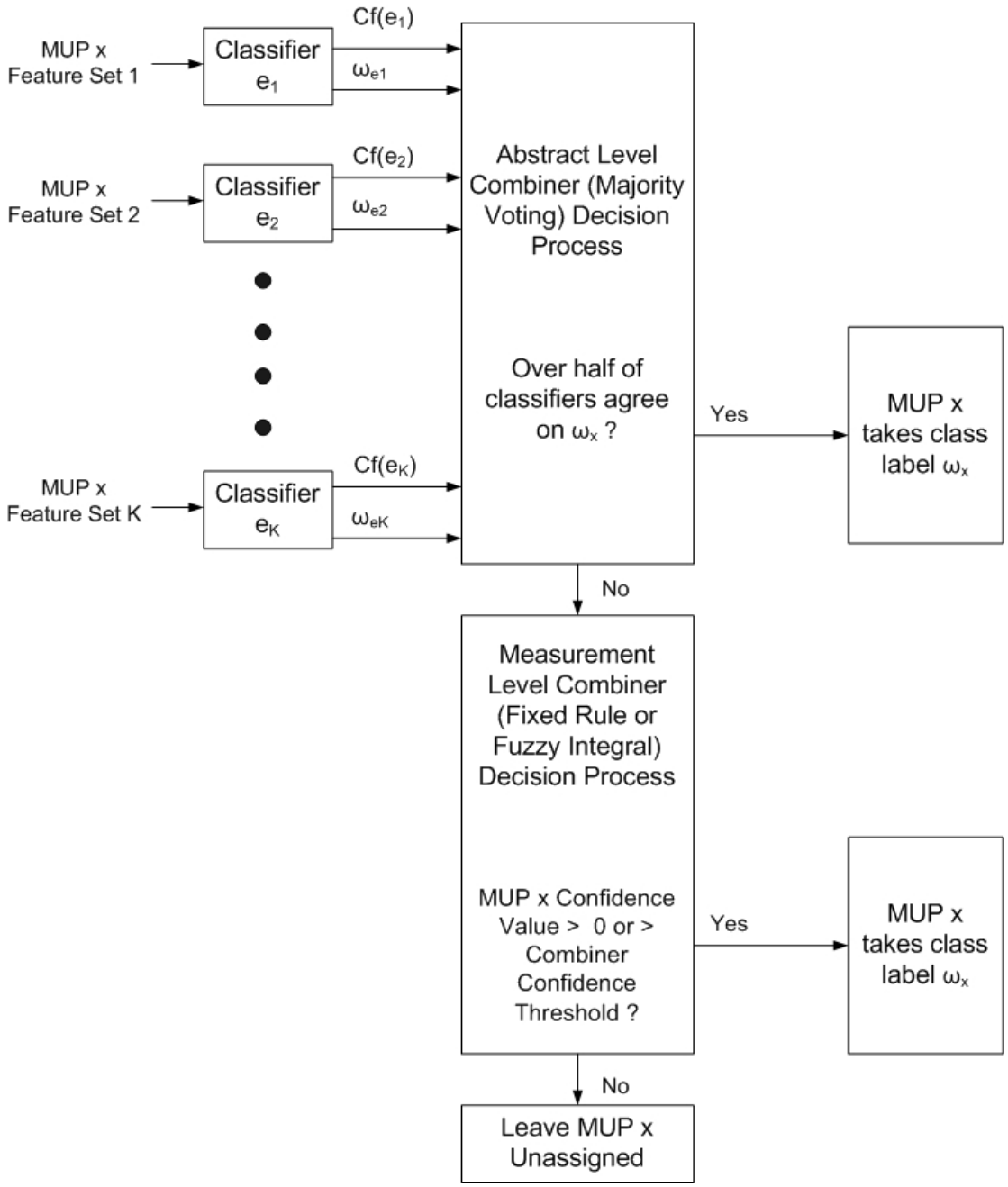

Figure 4.4: Hybrid classifier fusion scheme.

or Choquet fuzzy integral combiner. For each MUP, overall combined confidence values representing the degree of membership in each MUPT are determined and accordingly the MUP is assigned to the MUPT for which its determined overall combined confidence is the largest and if its above the specified combiner confidence threshold set for that MUPT. The MUPs whose overall combined confidence are greater than zero, or the specified combiner confidence threshold, are placed in the assigned MUPT thus forming a more complete set of MUPTs.

4.5.5 Diversity-Based Hybrid Classifier Fusion Scheme

The hybrid classifier fusion scheme described in Section 4.5.4 uses as a classifier ensemble a fixed set of classifiers and consequently both combiners act on the outputs of the same classifiers. The drawback of this scheme is apparent when following the *overproduce and choose* [36], [76] paradigm for the choice phase in the classifier fusion system design cycle, where there is a need to perform an exhaustive search for the best accuracy classifier ensemble. For example, if the pool of base classifiers contain 16 classifiers and the intention is to choose an ensemble of 6 base classifiers for fusion, so there is a need to compare the performance of $\binom{16}{6} = 8008$ ensembles. Therefore, to limit the computational complexity encountered we modified the hybrid classifier fusion scheme so that the candidate classifiers chosen for fusion by both combiners are based on a diversity measure.

The diversity-based hybrid classifier fusion scheme is a two-stage process and consists of two combiners with a pre-stage classifier selection module for each combiner. Once the pool of base classifiers has been constructed, it is not clear which subset of base classifiers when combined gives the best performance and as stated in Section 4.3 that the improvement in accuracy of classifier fusion system does not depend only on the method used for combining the base classifiers but also on the selection of classifiers used for combination. The ensemble candidate classifiers selected for combination are decided through assessing

the degree of agreement by exploiting a diversity measure for designing classifier teams. The kappa statistic measure, given in Sections 4.4 and 4.4.1, was used as the diversity measure to estimate the level of agreement between the base classifier outputs, i.e., to measure the degree of decision similarity between the base classifiers.

The MUP data set to be classified may contain MUPs with different complexities. There may be single MUPs each belonging to one MU, partially superimposed MUPs each representing the overlap of more than one MUP without their peaks being obscured, completely superimposed MUPs representing the overlap of more than one MUP and in which the peaks of the MUPs combine to make one large peak, and destructively superimposed MUPs representing the overlap of more than one MUP and in which the MUPs are superimposed in such a way that their out-of-phase peaks are summed together and cancel each other [25].

Based on the above MUP data set, the stage of designing the classifier ensemble mentioned in Sections 4.3, 4.4 for the diversity-based hybrid classifier fusion system is involved with the pre-stage classifier selection module for each combiner, such that the first combiner deals with combining the outputs of the classifier ensemble having the excellent level of agreement throughout all the MUPTs and the unassigned category using (4.9). This means that we want the first combiner to assign a large number of single MUPs correctly based on the highest level of consensus among the members of the ensemble and leave the classification of the more complex MUPs to a different classifier ensemble whose outputs to be combined by the second combiner. At the second combiner, the team's classifiers selection is based on assessing the classifier agreement considering only the unassigned category using (4.10) and choosing the ensemble having the least level of agreement. This is due to the fact that most classifiers face difficulties when classifying the more complex MUPs and don't assign them to any valid MUPTs and also due to the lack of consensus among the member of the ensemble.

The stage of designing the aggregation module mentioned in Sections 4.3, 4.5 for the

diversity-based hybrid classifier fusion system is involved with the choice of the combiners themselves. We chose the first combiner at the abstract level of classifier fusion and represented by the majority voting data independent aggregator. Whereas, we chose the second combiner at the measurement level of classifier fusion represented by the either one of the fixed combination rules or the fuzzy integral. Figure 4.5 shows the overall basis for the diversity-based hybrid classifier fusion scheme.

The diversity-based hybrid fusion scheme works as follows:

1. First stage: The ensemble candidate classifiers selected for combination by the first combiner are those having the maximum degree of agreement, i.e., having the maximum value of kappa statistics $\hat{\bar{\kappa}}$ evaluated using (4.9). The outputs of the ensemble candidate classifiers are presented to the majority voting combiner. If all the classifiers state a decision that a MUP is left unassigned, then there is no chance to re-assign that MUP to a valid MUPT and it stays unassigned. If over half of the classifiers assign a MUP to the same MUPT, then that MUP is allocated to that MUPT and no further assignment is processed. For these MUPs an overall confidence value is calculated by averaging the confidence values given by the ensemble classifiers who contributed in the decision of assigning the MUP. In all other situations, i.e., when half or less than half of the classifiers specify a decision for a MUP to be assigned to the same MUPT, the measurement level combination scheme is used in the second stage to specify to which MUPT the MUP should be assigned based on which MUPT has the largest combined confidence value. From the first stage, a set of incomplete MUPTs are generated missing those MUPs that need to be assigned to a valid MUPT in the second stage.

2. Second stage: This stage is used for those MUPs for which only half or less than half of the ensemble classifiers in the first stage specify a decision for a MUP to be assigned to the same MUPT. The ensemble candidate classifiers selected for combination at the second combiner are those having a minimum degree of agreement considering

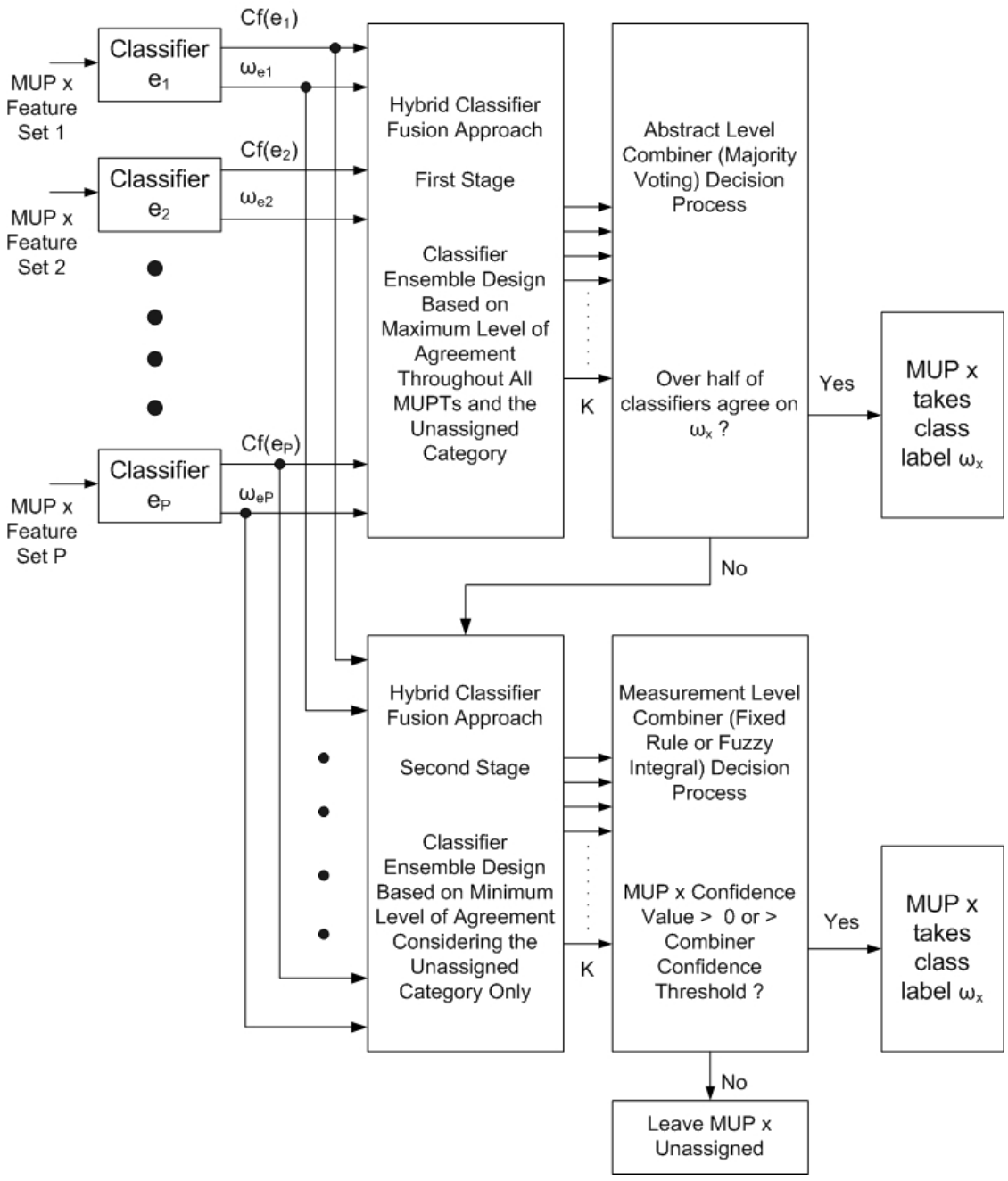

Figure 4.5: Diversity-based hybrid classifier fusion scheme.

only the unassigned category, i.e., having the minimum value of kappa statistics $\hat{\kappa}_i$ evaluated using (4.10) for $i = M + 1$. The outputs of the ensemble classifiers are presented to one of the fixed rule combiners or one of the trainable combiners represented by Sugeno or Choquet fuzzy integral combiner. For each MUP, overall combined confidence values representing the degree of membership in each MUPT are determined and accordingly the MUP is assigned to the MUPT for which its determined overall combined confidence is the largest and if its above the specified combiner confidence threshold set for that MUPT. The MUPs whose overall combined confidence are greater than zero, or the specified combiner confidence threshold, are placed in the assigned MUPT thus forming a more complete set of MUPTs.

4.5.6 The Adaptive Classifier Fusion Approach

To reduce the number of erroneous assignments in the generated MUPT by the one level classifier fusion scheme, the hybrid, and the diversity-based hybrid classifier fusion approach, we adopt the same adaptive setting of train assignment threshold used for base classifiers and described in Sections 3.3.1 and 3.3.2.

As stated in Section 3.3.2, the adaptive nature of MUP classification is related to the adjustment of the minimal assignment threshold for each MUPT based on firing pattern statistics. In the classifier fusion approach we start with a very low assignment threshold and test all MUPTs against firing pattern constraints. The assignment threshold of a MUPT might be increased to exclude MUPs causing firing pattern inconsistencies. The occurrence of a significant number of MUP classification errors is detected by the use of the firing pattern consistency statistics as presented in Section 3.2.1.

MUP classification using the classifier fusion approach is based on the outputs of the base classifiers and it does not take into consideration MU firing patterns. Therefore, following the generation of MUPTs by the classifier fusion system, the MUPTs should be

tested for any firing pattern inconsistencies and if, based on firing pattern statistics given in Section 3.2.1, it is expected that a train has too many erroneous assignments, its minimal assignment threshold is increased or otherwise it is kept constant. This firing pattern analysis allows the algorithm to modify the assignment of a MUP for each train individually based on an expectation of the number of erroneous assignments. Trains to which MUPs can be confidently assigned will have a lower minimal assignment threshold and have a higher MUP identification (ID) rate. Alternatively, trains to which MUPs cannot be confidently assigned, possibly because of shape similarity with the MUPs of another train or MUPT shape and firing pattern variability, will have a higher minimal assignment threshold and have a lower MUP identification rate. In both cases, the number of errors expected will be approximately constant. Therefore, over all the trains, a maximum number of MUPs should be assigned while maintaining an acceptable MUP assignment error rate. The adjustment of the minimal assignment threshold of each MUPT based on the pattern of MUP occurrences in the train represents a passive use of firing pattern information. Specific class assignments are not actively determined by the firing pattern information. Rather, the firing pattern information is used to allow or deny MUP assignments in general.

Unlike the adaptive supervised classification task employed by base classifiers, during classifier fusion the adaptive MUP classification consists of only one stage and it is based on the passive use of firing pattern information to remove possible erroneous classifications. Figure 4.6 shows the steps involved in this process. Each candidate MUP j is assigned to the train that has the highest confidence of belonging to it expressed in terms of the maximum assignment value and if it is higher than the minimum assignment threshold of that train.

Once the set of MUPTs are created by the classifier fusion system, the firing pattern consistency statistics for each MUPT with $N_{sufficient}$ or more IDIs are calculated. A small percent of those MUPs causing IDI inconsistencies are allowed. The firing pattern consistency statistics of each MUPT with $N_{sufficient}$ or more IDIs are then re-calculated and

checked against the following constraints:

Percentage number of IDI inconsistencies $\leq 2.5\%$.

Absolute value of $CV_l \leq 0.35$, representing an upper bound of physiological expectation.

Ratio of CV_l to $CV \leq 1.25$, empirically determined value.

Lower IDI ratio ≤ 0.175, empirically determined value.

If a MUPT meets all the above constraints, it keeps its MUPs and its minimum assignment threshold is unchanged. Otherwise, the minimum assignment threshold for the train is increased by $\Delta_{increase}$ and its MUPs confidence values are checked against the new value. Those MUPs causing IDI inconsistencies in the firing pattern of a train and MUPs with confidence values less than the minimum assignment threshold are designated unassigned and removed from the train. This process is repeated until all the imposed firing pattern constraints for all generated MUPTs are satisfied. A MUPT's minimum assignment threshold is not increased above an extreme value or if it has fewer than a minimum number of MUPs $N_{sufficient}$. In all the experiments concerning the evaluation of classifier fusion systems, the required number of MUPs in a MUPT to make assignment threshold changes, $N_{sufficient}$, is taken to be 50 in order to enable the EFE algorithm estimates the MUPT mean and standard deviation with reasonable error.

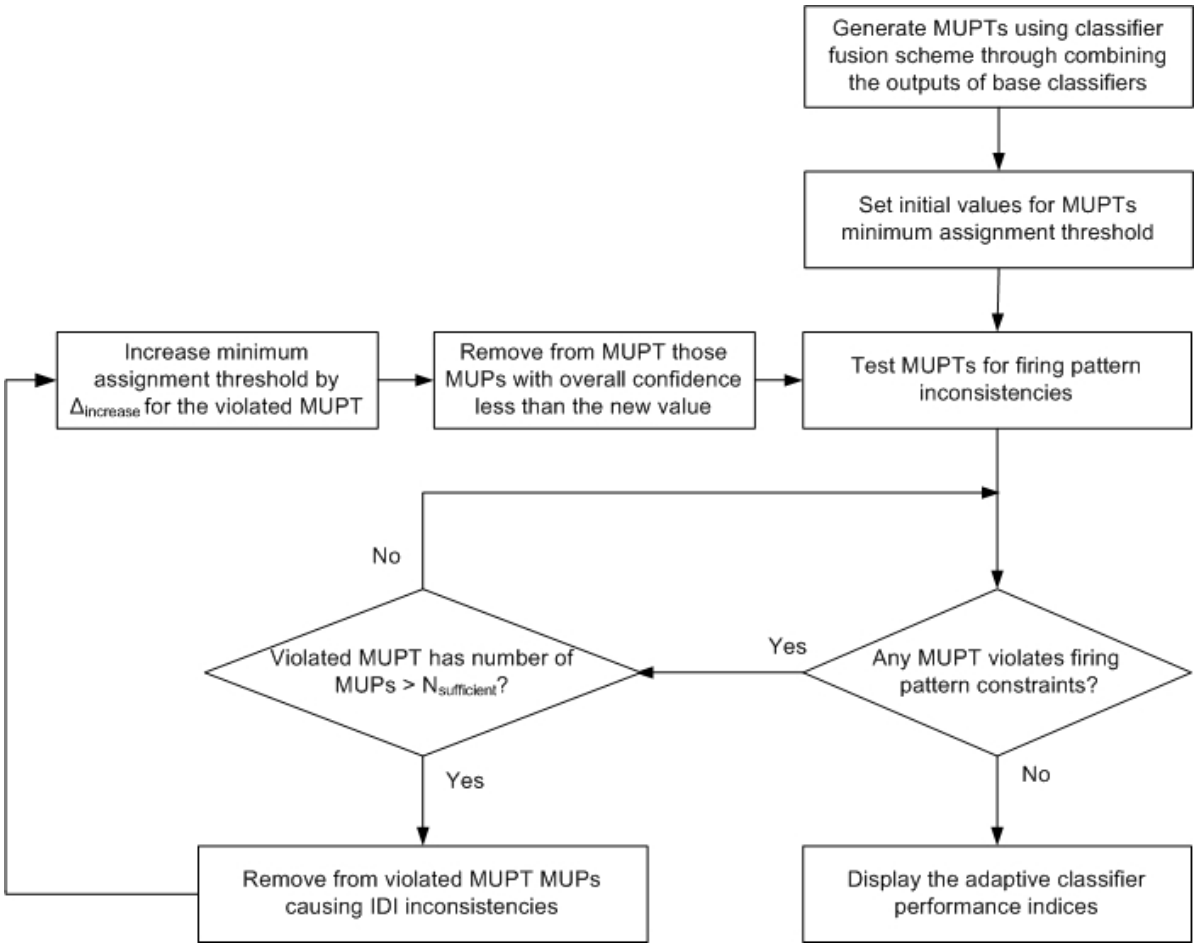

Figure 4.6: Flowchart of adaptive classification using a classifier fusion scheme.

Chapter 5

Results and Comparative Study

5.1 Introduction

The effectiveness of a classifier fusion system as applied to the EMG signal decomposition process is demonstrated through a series of simulations and applications to real EMG signals.

The classification task completed by the base classifiers uses different domain feature sets: MUP time-domain and wavelet-domain features. The performance of each individual classifier and classifier fusion system was evaluated in terms of the numbers of assigned and rejected MUP waveforms and the number of correctly and erroneously classified MUP waveforms and from which a set of related performance indices was determined. Also a confusion matrix for each classifier was computed.

The information provided by the individual classifier confusion matrix could be regarded as a priori knowledge for managing the uncertainty in classifying a MUP. From which we could collect output information in the form of a posteriori probabilities necessary to combine classifier decisions at the measurement level.

This chapter describes the evaluation process for the base classifiers and the classifier fusion system and presents the methods employed for the evaluation. The EMG signal

data sets used, the feature extraction and base classifier seeding methods, and the classifier performance indices are described. Also provided are results across the sets of simulated and real EMG signals in terms of the performance indices of the individual different types of base classifiers; and the multiple classifier fusion system with different fusion schemes and approaches. Finally a summary of experimental results and a comparative study discussion is provided.

5.2 Classifier Evaluation

The testing process and evaluation of performance of the base classifiers and the classifier fusion system requires a reference decomposition result. For this purpose, different approaches can be followed [29], [30], [66]. Two of these approaches are:

1. Using simulated EMG signals with specific properties as a reference,

2. Using real EMG signals decomposed manually by a human expert as a reference.

For testing and evaluating the classifiers, we followed the two approaches mentioned above.

We used simulated EMG signals generated from an EMG signal simulator based on a physiologically and morphologically accurate muscle model [42]. We also used a set of real EMG signals of different complexities that were detected during slight to moderate levels of contraction. The methods employed for the evaluation process are described in the following sections.

5.2.1 MUP Detection and Representation

The first task in EMG signal decomposition is the segmentation and MUP detection task. It is concerned with locating the main positive peaks or spikes found in an EMG signal. The detected spikes or MUPs should have rapid rising edges, which indicates that the electrode is close to active muscle fibres. Motor units that were active during signal acquisition

generate these MUPs. Conversely, MUPs that have slow rising edges and small amplitude were generated from motor units that have fibres far away from the electrode.

The EMG signal is divided into segments of possible MUP waveforms and searching for time intervals containing these MUPs comprises the MUP detection operation. A segment can either contain one MUP or superimposed MUPs (compound segments). Time intervals with low energy are without MUPs and represent the signal baseline. The detected spikes within windows of sampled raw data or its first and second-order discrete derivatives form the MUP waveforms. A window of 80 sample points, representing MUP intervals of 2.56 ms at a sampling rate of 31.25 kHz, formed a MUP pattern feature vector. The collection of feature vectors form the feature space data set necessary for subsequent pattern recognition operations.

The other feature space data sets studied were comprised from the wavelet-domain coefficients and extracted by applying the discrete-time wavelet transform (DTWT) to MUP raw data or its first-order and second-order discrete derivatives. The use of wavelets for feature extraction gives the possibility of looking at several different representations of the MUP pattern data by using different types of wavelet filters and different wavelet decomposition scale levels.

From the time-domain and wavelet-domain extracted features we applied a soft thresholding de-noising procedure to get a de-noised set of features and fed them to the classifiers for comparison purposes.

5.2.2 Discrete Derivative Features

Discrete derivative features deal with the phase space of the data [35]. For each EMG signal investigated, we extracted from the MUP raw data feature space the first-order discrete derivative features by applying a low-pass differentiation filter to accentuate the MUP's rapidly rising edges (high frequency components) and shorten the duration of MUPs, which reduces their temporal overlap and therefore reduces the number of superimposed wave-

forms. This filtering also suppresses low frequency background noise. The effect is to transform the MUPs into narrow spikes that stand out clearly from a flat baseline.

Denoting the discrete samples of the raw data as $x[i]$ and the first-order discrete derivative data as $y^{(1)}[i]$, then to perform first-order differentiation we used the filter given by (5.1) and (5.2) [68], [107]:

$$\begin{array}{c} y^{(1)}[i] = (\frac{1}{2}) \sum_{k=1}^{W} C_k \ (x[i+k] - x[i-k]) \\ with \qquad \sum_{k=1}^{W} k.\ C_k = 1 \end{array} \tag{5.1}$$

where, W is the window width for adjusting the cut-off frequency. Taking the window width $W = 3$, the optimum coefficients C_k are $C_1 = C_2 = 0,\ C_3 = \frac{1}{3}$. For these values, (5.1) becomes:

$$y^{(1)}[i] = (\frac{1}{2})(\frac{1}{3}) \ (x[i+3] - x[i-3]) \tag{5.2}$$

Denoting the second-order discrete derivative data as $y^{(2)}[i]$, then to perform second-order differentiation we used the filter given by (5.3) and (5.4) [68], [107]:

$$\begin{array}{c} y^{(2)}[i] = (\frac{1}{2}) \sum_{k=1}^{W} C_k \ (x[i+k] + x[i-k]) \\ with \qquad \sum_{k=1}^{W} k^2.\ C_k = 2 \end{array} \tag{5.3}$$

Taking the window width $W = 3$ and optimum coefficients C_k, (5.3) becomes:

$$y^{(2)}[i] = (\frac{1}{2})(\frac{1}{3}) \ (x[i-3] - x[i] - x[i+3] + x[i+6]) \tag{5.4}$$

5.2.3 Wavelet Transform Coefficient Features

The wavelet transform (WT) provides a linear two-dimensional time-frequency representation by decomposing a signal into different frequency components, and then represents each component with a resolution matched to its scale [18]. As the EMG signals to be decomposed are in discrete-time form, we only investigated the discrete-time wavelet transform (DTWT) with orthogonal wavelet bases.

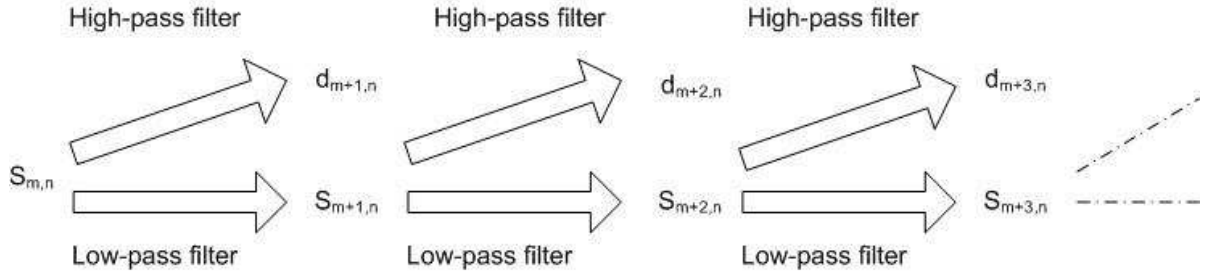

Figure 5.1: Wavelet decomposition algorithm of the fast wavelet transform.

With the goal of extracting meaningful features combined with the intrinsic nonstationary time behavior nature of EMG signals, time-frequency representations are highly desirable because they have the ability to localize the changes in the statistics of nonstationary EMG signals. The time-frequency analysis of a signal is processed by decomposing the signal on dilated and translated versions of a mother wavelet function. The time-frequency window depends on the dilation factor of the wavelet; it is long at low frequencies and short at high frequencies.

The wavelet-domain features are extracted by taking the DTWT of the EMG signal MUP patterns. The DTWT is computed using the wavelet decomposition algorithm of the fast wavelet transform (FWT) through multiresolution analysis and is implemented using a filter bank structure consisting of only the analysis bank. At each level of the transform, two filters are used for processing the data through a low-pass and high-pass filter. The low-pass filtered data are known as the approximation wavelet coefficients. The high-pass filtered data are known as the detail wavelet coefficients. The approximation wavelet coefficients can be sent again as input data to compute the next level approximation and detail wavelet coefficients; thus we can decompose the signal into its different frequency components at different scale levels.

The orthogonal wavelet coefficients of MUP samples were computed from the decomposition relation depicted in Figure 5.1. The approximation $s_{m,n}$ and detail $d_{m,n}$ coefficients were calculated with a cascade of discrete convolutions and subsamplings.

The discrete-time MUP samples, $x[i]$, with $i = 0, 1, ..., N-1$ and N being the MUP sample size were input directly as the approximation coefficients at scale m_0, i.e., $s_{0,n} = x[n]$ where at scale $m = 0$, both the coefficient location index n and the MUP waveform discretization index i have the same range (0 *to* $N-1$) and were equal to each other. Then the multiresolution analysis began by computing the approximation and detail coefficients at all scales larger than m_0 using the following formulas:

$$s_{m+1,n} = \frac{1}{\sqrt{2}} \sum_k g_k .\ s_{m,2n+k} = \frac{1}{\sqrt{2}} \sum_k g_{k-2n} .\ s_{m,k} \tag{5.5}$$

$$d_{m+1,n} = \frac{1}{\sqrt{2}} \sum_k h_k .\ s_{m,2n+k} = \frac{1}{\sqrt{2}} \sum_k h_{k-2n} .\ s_{m,k} \tag{5.6}$$

Hence, using (5.5) and (5.6), we generated the approximation and detail coefficients at a specific scale from the approximation coefficients at the previous scale. The vector containing the sequences $(\frac{1}{\sqrt{2}})g_k$ is the low-pass filter and $(\frac{1}{\sqrt{2}})h_k$ is the high-pass filter. Corresponding to each type of wavelet, there is a corresponding sequence for the low-pass and high-pass filters.

With a discrete-time MUP pattern of sample size N and sequence length of each of the low-pass and high-pass filters being $2K$, the detail d_1 and approximation s_1 coefficients sequence length was:

$$floor(\frac{N-1}{2} + K) \tag{5.7}$$

where $floor(\frac{N-1}{2})$ rounds $\frac{N-1}{2}$ to the nearest integer.

The wavelet-domain full MUP pattern feature vector after the full wavelet decomposition with M scale levels was the concatenation of the detail coefficients at all scale levels and the approximation coefficients at the last scale level and had the form $V^{(M)} = (s_M, d_M, d_{M-1}, ..., d_2, d_1)$. The wavelet-domain MUP feature vector used for classification purposes was a subset of the full feature vector containing only the detail coefficients at the high scale levels. For $M = 6$ scale levels, we noticed that taking only the detail coefficients at levels 4, 5, and 6 gave better classification performance.

5.2.4 Wavelet De-noising Features

We applied an optimal de-noising procedure called soft-thresholding [21] using discrete wavelet analysis to reduce the interference of white noise. It performed thresholding in the wavelet domain, and it involved keeping only the portion of the details that exceeded a certain limit while almost entirely suppressing the noise and then reconstructing the signal by computing the inverse discrete wavelet transform.

Using wavelets to remove noise from a signal requires identifying which component or components contain the noise, and then reconstructing the signal without those components. When decomposing the MUP waveform using wavelet analysis, we noted that successive approximations became less and less noisy as more and more high-frequency information was filtered out of the signal. Therefore we decomposed the signal to a level at which the approximation signal was somehow clean as compared to the original signal. In discarding all the high-frequency information, we lost many of the original signal's sharpest features. But with soft-thresholding de-noising, the reconstructions had two properties [20]: the noise was almost entirely suppressed, and significant features sharp in the original signal remained sharp after reconstruction.

5.2.5 MUP Alignment

The simulated EMG signal data sets used in this thesis for evaluating the base classifiers contain signals with different MUP shape and/or IDI variability. In order to reduce the effect of these variabilities and to accurately compare MUPs for the classification task, the MUPs were first aligned.

We chose the *maximum-correlation alignment* algorithm [68], [69] for alignment purposes to extract the integer offset that was applied in discrete time to the candidate MUPs for alignment. For each train a typical MUP was taken to be the mean of all the MUPs belonging to it.

5.2.6 Seeding the Classifiers

The supervised classification task of base classifiers relies on a labelled reference set that contains correctly classified MUPs. The reference set of MUPs are those MUPs used to seed base classifiers. The developed system uses the following methods to specify the seeding MUP data:

1. Clustering Data: The clustering operation in the EMG signal decomposition provides the supervised classification stage with initial knowledge about the number of active motor units (clusters) and the MUPs assigned to each determined motor unit. The clustering operation is performed on a portion of the EMG signal that has the highest activity, i.e., with maximum intensity of MUPs. Clustering operations are based on the MUP shape information and should take into consideration the firing pattern information. The *shape and temporal based clustering* (STBC) algorithm [81], [100] takes the constraint of incorporating the firing pattern information during clustering and is used for providing the base classifiers with the labelled reference set.

2. Sequential Assignment Data: A reference set of correctly classified MUPs is chosen from simulator generated information or manually decomposed real data. For each MUPT, beginning from its head, we choose as the reference set a number of isolated MUPs that are apart by more than 3 ms to avoid choosing superimposed MUPs.

3. High Certainty Data: This is data that are generated by a seeding classifier, which is responsible for accurately calculating MUPT templates. The seeding classifier is a single-pass Certainty classifier that selects from the set of detected MUPs: a number of MUPs from each train for accurately calculating the train templates. The single-pass classifier itself was seeded by a set of MUPs satisfying Forgy's criterion [48]. This is equivalent to selecting the core class members. Let $d(m_i, s_j)$ be the distance from MUP m_i to MUPT j centre s_j and $\overline{d}_i$ be the average distance from MUP m_i to

all M MUPT centres:

$$\overline{d}_i = \frac{1}{M} \sum_{j=1}^{M} d(m_i, s_j) \tag{5.8}$$

Each MUPT center is specified to be the mean of all the MUPs belonging to it and as determined by the simulator or human expert. Then a MUP was added to the class core if:

$$|d(m_i, s_q) - \overline{d}_i| \leq \ T_F \ |\overline{d}_i| \tag{5.9}$$

where the qth class center s_q is the class center closest to MUP m_i and T_F is a threshold between zero and one. For larger values of T_F, more seeding MUPs are added to the class core. The threshold T_F was set automatically. For each MUPT, the MUPs that satisfy equation (5.9) were added to the class core using an initial threshold value of 0.05, which was increased incrementally until the desired number of MUPs in the class core were selected. The class core MUPs were then used to calculate the single-pass classifier templates. Once the classification pass ended, a desired number of MUPs having the highest shape certainty were selected from each MUPT and those MUPs would be the labelled reference set. The highest shape certainty decision for assigning a MUP to a MUPT is based on the value for which the multiplicative combination of the normalized absolute shape certainty C_{ND}, given by (3.4), and the relative shape certainty C_{RD}, given by (3.6), is the highest.

After specifying each train labelled reference set, we made sure that we choose isolated MUPs that were apart by more than 3 ms to avoid choosing superimposed MUPs.

When using the clustering data or the sequential assignment data for seeding a base classifier, the classifier was able to track the non-stationarity of the MUPs waveform shape by updating the labelled reference set once the MUP to be assigned has a shape confidence higher than an updating confidence threshold.

In case of classifiers based on fuzzy k-NN rule, the MUPs reference set membership function estimation may be either a crisp labelling or soft labelling using (3.20) or (3.21)

and based on experimentation with several EMG signals we found that (3.21) gave better results. The MUPs reference set was used for selecting the nearest neighbours of each candidate MUP to be classified. The reference set update was performed by replacing the MUP to be assigned having the high assertion with a MUP having a less assertion in the same MUPT.

For the certainty-based classifiers or matched-template filter classifiers, the reference set of MUPs was used to calculate the template of each train. The reference set update was performed by the following updating rule:

$$s_{i+1} = \frac{s_i + Cf_c \;.\; m_c}{1 + Cf_c} \tag{5.10}$$

Where s is the moving-average template vector, m_c is the classified MUP feature vector whose confidence Cf_c exceeds the updating confidence threshold [75], [99].

5.2.7 Classifier Performance Indices

EMG signal decomposition is performed to obtain information about the individual MUPs that comprise the signal as well as the individual activation patterns of the MUs that generate them. As such, a classifier used for EMG signal decomposition must assign as many detected MUPs as possible while at the same time limiting the number of classification errors so that accurate MUP templates, or accurate nearest neighbors in the case of the fuzzy k-NN classifier, and MU activation patterns can be determined. These conflicting constraints require a classifier to leave MUPs, which cannot be assigned with sufficient confidence, unassigned.

The classification performance of the base classifiers and the classifier fusion system were evaluated and compared in terms of their assignment rate, error rate, and correct classification rate performance indices.

The assignment rate $A_r\%$ was defined as the ratio of the total number of assigned MUPs, which is equal to the total number of MUPs detected minus the number of MUPs

unassigned, to the total number of MUPs detected:

$$A_r\% = \frac{number\ of\ MUPs\ assigned}{total\ number\ of\ MUPs\ detected} * 100 \tag{5.11}$$

The error rate $E_r\%$ was defined as the ratio of the number of MUPs erroneously classified to any valid MUPT to the number of MUPs assigned:

$$E_r\% = \frac{number\ of\ MUPs\ erroneously\ classified}{number\ of\ MUPs\ assigned} * 100 \tag{5.12}$$

The correct classification rate $CC_r\%$ was defined as the ratio of the number of correctly classified MUPs, which is equal to the number of MUPs assigned minus the number of MUPs erroneously classified, to the total number of MUPs detected:

$$CC_r\% = \frac{number\ of\ MUPs\ correctly\ classified}{total\ number\ of\ MUPs\ detected} * 100 \tag{5.13}$$

The correct classification rate CC_r can be written in terms of the assignment rate A_r and the error rate E_r using the relation $CC_r = A_r(1 - E_r)$.

Classifier performances are compared in terms of the above set of performance indices such that the classifier with the *better performance* is the one with the highest correct classification rate $CC_r\%$ and lowest error rate $E_r\%$. In situations when the highest correct classification rate $CC_r\%$ and lowest error rate $E_r\%$ does not judge the differentiation between two classifiers, we take the difference between the correct classification rate $CC_r\%$ and error rate $E_r\%$ for each classifier and consider the classifier with the higher difference as the one with the better performance.

5.2.8 EMG Signal Data Sets Used

The simulator enables us to generate EMG signals of different complexities with knowledge of the signal intensity represented by the average number of MUP patterns per second (pps), the numbers of MUPTs, and which motor unit created each MUP pattern. Furthermore, the amount of MUP shape variability represented by jitter and/or IDI variability can be

adjusted. Jitter represents the standard deviation of the initiation times of each muscle fibre potential [42]. In normal subjects, jitter values of about 25 μs are expected. The goal of these simulations was to produce data to compare the performance of the base classifiers and the classifier fusion system when fed with time-domain and wavelet-domain first-order discrete derivative features.

Two sets of simulated EMG signals: independent and related, each of 10 seconds length, were generated by the simulator. Twenty independent EMG signals with different levels of intensity, ranging from 30.5 pps to 127.5 pps, and each having unique MUPTs and MUP distributions, with jitter values ranging from 25 μs to 100 μs, but with equal degree of IDI variability ($CV = 0.15$) were created. The characteristics of these signals are listed in Table 5.1. These data allowed performance related to signal intensity, number of trains and MUP shape variability to be evaluated.

Three groups of related EMG signals were also created:

1. The first group has 9 signals. Each signal in this group has 9 MUPTs and represents the contributions from the same 9 MUs. The average signal intensity across these 9 signals was 95 pps, which corresponds to an average MU firing rate of 10.6 pps (or a mean IDI of 95 ms). However, each EMG signal within the group was generated by imposing different amounts of biological variability of the MUP shapes from discharge to discharge within a MUPT using jitter values of 50, 75, and 150 μs, respectively, and imposing different IDI variability using CV values of 0.1, 0.15, 0.2, 0.3, and 0.45, respectively, (0.10 to 0.15 is expected for healthy subjects). Therefore, across the 9 signals of the group, the same MUs contribute to the composite EMG signal and each MU contributes the same MUPs on average, however, the variability of the MUPs in each train will increase with the jitter value. Likewise, the mean IDI of each train is the same across the signals of the group but the variability of the IDIs in each variation increases with the CV value.

2. The second group has 4 signals each has 6 MUPTs. The average signal intensity

Table 5.1: Characteristics of the independent simulated EMG signals.

EMG signal	Intensity (pps)	No. of MUPTs	Jitter (μs)
1	30.5	3	100
2	35.3	4	50
3	41.8	5	100
4	45.6	4	50
5	54.0	6	50
6	59.4	7	100
7	61.4	6	75
8	68.2	7	50
9	70.7	7	100
10	79.3	8	25
11	82.5	8	100
12	85.2	9	50
13	91.7	7	50
14	97.5	10	100
15	105.2	9	50
16	109.5	9	75
17	116.5	8	75
18	119.4	10	75
19	120.6	11	50
20	127.5	11	50

across these 4 signals was 62 pps, which corresponds to an average MU firing rate of 10.3 pps (or a mean IDI of 97 ms). Three of the signals in this group were generated with equally variably MUPs by using a fixed 150 μs jitter value but each had a different amount of variability in the firing patterns by using IDI CV values of 0.1, 0.15, and 0.3, respectively. A fourth member of this group was created with an IDI CV value of 0.15 and a 50 μs jitter value.

3. The third group has 3 signals each with 10 MUPTs. The average signal intensity across these 3 signals was 134 pps, which corresponds to an average MU firing rate of 13.4 pps (or a mean IDI of 75 ms). The EMG signals in this group were generated with different biological variability of the MUP shapes by using jitter values of 50, 75, and 150 μs, respectively, and a fixed IDI CV of 0.3.

The characteristics of the groups of related signals are listed in Table 5.2. These three groups of data allow the performance of the classifiers to be evaluated with respect to shape and firing pattern variability jointly or independently and to assess performance consistency across a number of detection sites and signal intensity scenarios.

Five real EMG signals of different complexities, whose characteristics are listed in Table 5.3, were also analyzed for the evaluation process. The real signals were detected during slight to moderate levels of contraction corresponding to approximately 10 to 25% of maximum voluntary contraction (MVC), of the extensor digitorum communis (EDC) muscle of normal subjects.

The real EMG signals have been decomposed manually by an experienced operator using a computer-based graphical display algorithm. The manual decomposition results were assumed to be the reference and were compared with those obtained automatically by the classifiers.

Table 5.2: Characteristics of the groups of related simulated EMG signals.

EMG signal	Intensity (pps)	No. of MUPTs	Jitter (μs)	CV
1-1	94.4	9	50	0.15
1-2	94.4	9	75	0.15
1-3	94.4	9	150	0.15
1-4	95.5	9	50	0.1
1-5	96.0	9	50	0.2
1-6	95.4	9	50	0.3
1-7	92.9	9	50	0.45
1-8	96.1	9	75	0.3
1-9	94.4	9	150	0.3
2-1	62.4	6	50	0.15
2-2	62.4	6	150	0.1
2-3	61.5	6	150	0.15
2-4	62.7	6	150	0.3
3-1	134.4	10	50	0.3
3-2	135.0	10	75	0.3
3-3	133.6	10	150	0.3

Table 5.3: Characteristics of the real EMG signals.

EMG signal	Intensity (pps)	No. of MUPTs
1	76.3	6
2	72.3	8
3	96.3	8
4	63.0	6
5	80.4	6

5.3 Data Analysis Results

This section reports classification performance results of the assigned, rejected, correctly classified, and erroneously classified MUP waveforms and presents them in terms of the assignment rate A_r, error rate E_r, and correct classification rate CC_r defined in Section 5.2.7. The results obtained using the base classifiers and the classifier fusion schemes and approaches are presented in this section. With respect to the classifier fusion system, we present, for the purpose of comparison, results of the one level classifier combiner and the hybrid classifier fusion system. The performance with respect to each signal in the EMG signal data set given in Section 5.2.8, the mean performance and the mean absolute deviation (MAD) of the performance are reported.

In each of the EMG signals considered, superimposed waveforms were included in the set of MUPs to be assigned. Most of these superimposed MUPs were left unassigned by each of the classifiers. If a superimposed MUP was assigned, it was considered a correct classification if it was assigned to one of the MUPTs whose MUPs comprised the superimposed waveform. Otherwise, it was considered a false classification. The EFE algorithm when estimating the mean IDI and standard deviation successfully excluded long IDIs caused by missed classifications whether due to superimposed MUPs or other causes.

5.3.1 Base Classifier Results

Results, in terms of assignment rate A_r, error rate E_r, and correct classification rate CC_r, obtained using the ACC, AFNNC and the AMTFC (two variants: the ANCCC and the ApCC) base classifiers across the two sets of simulated EMG signals are presented in Tables 5.4 and 5.5, 5.7 and 5.8, 5.10 and 5.11, 5.13 and 5.14, respectively. Tables 5.6, 5.9, 5.12, and 5.15 show the classifiers performance results using the real signals. The performance with respect to each signal, the mean performance and the mean absolute deviation (MAD)

of the performance are reported.

The base classifier results presented in this section were obtained for four classifiers from each kind, i.e., for ACC classifiers e_1, e_2, e_3, e_4, for AFNNC classifiers e_5, e_6, e_7, e_8, for ANCCC classifiers $e_9, e_{10}, e_{11}, e_{12}$, and for ApCC classifiers $e_{13}, e_{14}, e_{15}, e_{16}$. Classifiers e_1, e_5, e_9, e_{13} were fed with time-domain first-order discrete derivative features and using high-certainty MUPs for seeding. Classifiers e_2, e_6, e_{10}, e_{14} were fed with time-domain first-order discrete derivative features and using MUPs with sequential assignment for seeding. Classifiers e_3, e_7, e_{11}, e_{15} were fed with wavelet-domain first-order discrete derivative features and using the highest shape certainty MUPs for seeding. Classifiers e_4, e_8, e_{12}, e_{16} were fed with wavelet-domain first-order discrete derivative features and using MUPs with sequential assignment for seeding. Note that, throughout the results presentations, we follow the previous labeling for base classifier identification.

The seeded data were taken from the reference data supplied by the simulator and the manually decomposed data. The selected reference set was used to calculate initial train templates for the ACC, ANCCC, and ApCC, and to establish core membership values for the AFNNC.

The wavelet-domain features were extracted through multiresolution analysis with Daubechies 4 wavelet filters and 6 scale levels and then forming the feature vectors using only the detail coefficients at levels 4, 5, and 6.

For the adaptive fuzzy k-NN AFNNC classifier, the presented results were produced using the following settings: number of nearest neighbors $k = 5$; Keller soft labelling given by (3.21) and Euclidean distance measure. These settings for the AFNNC were empirically found to perform better based on experimentation with several EMG signals.

Table 5.4: Performance of Adaptive Certainty Classifier using the independent simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_1 **- High Certainty**			e_2 **- Sequential Assignment**			e_3 **- High Certainty**			e_4 **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	91.2	0.7	90.6	91.2	0.7	90.6	95.3	0.0	95.3	94.9	0.4	94.6
2	94.8	0.3	94.5	97.1	0.3	96.8	95.1	0.3	94.8	97.1	0.3	96.8
3	87.1	1.4	85.9	90.0	0.8	89.3	86.9	1.7	85.4	90.8	0.8	90.0
4	88.6	1.8	87.0	89.7	1.5	88.3	91.3	3.4	88.1	92.2	1.9	90.4
5	94.8	0.4	94.4	96.3	0.4	96.0	95.4	0.8	94.6	96.3	0.4	96.0
6	91.8	0.6	91.2	93.5	0.6	93.0	91.2	0.6	90.7	92.6	0.4	92.3
7	81.8	1.5	80.6	82.0	1.5	80.8	85.8	1.8	84.3	87.6	2.0	85.8
8	92.2	1.2	91.1	93.6	0.7	93.0	92.0	1.2	90.9	93.0	0.7	92.3
9	86.5	0.7	85.9	89.9	0.7	89.3	88.1	1.9	86.5	90.0	1.2	89.0
10	88.4	0.6	87.8	90.3	1.0	89.3	89.2	1.2	88.1	91.9	1.0	90.9
11	86.6	1.2	85.6	86.1	4.2	82.5	89.2	2.2	87.2	89.7	4.5	85.7
12	86.6	1.7	85.1	87.6	1.6	86.2	87.8	2.6	85.6	88.1	2.4	85.9
13	78.7	2.6	76.6	80.6	2.1	78.9	79.1	2.4	77.2	78.2	1.4	77.1
14	85.6	2.1	83.8	89.0	1.6	87.5	87.6	3.6	84.5	90.8	2.5	88.5
15	79.9	2.6	77.8	81.4	2.1	79.7	82.3	2.9	79.9	85.5	2.2	83.6
16	82.2	1.6	80.9	83.4	1.2	82.5	84.9	2.8	82.5	86.3	2.7	84.0
17	82.1	2.2	80.3	81.4	2.0	79.8	84.7	2.7	82.5	83.6	2.4	81.6
18	70.6	6.3	66.2	75.4	5.1	71.6	72.5	10.5	64.9	78.0	8.7	71.2
19	72.8	3.7	70.1	76.4	3.5	73.8	75.9	4.6	72.4	79.9	4.5	76.3
20	80.4	2.3	78.8	83.1	1.5	81.9	83.4	2.3	81.5	85.8	2.3	83.9
Mean	85.1	1.8	83.7	86.9	1.6	85.5	86.9	2.5	84.8	88.6	2.1	86.8
MAD	5.2	0.9	5.8	5.2	0.9	5.8	4.6	1.3	5.4	4.5	1.3	5.3

Table 5.5: Performance of Adaptive Certainty Classifier using the groups of related simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_1 **- High Certainty**			e_2 **- Sequential Assignment**			e_3 **- High Certainty**			e_4 **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	89.9	2.3	87.8	88.3	1.3	87.1	90.7	2.9	88.0	92.5	1.9	90.8
1-2	85.2	2.2	83.4	90.4	1.7	88.9	90.0	2.6	87.7	90.7	2.4	88.5
1-3	84.4	2.5	82.3	85.3	4.0	81.8	87.9	3.8	84.6	86.7	4.9	82.4
1-4	87.5	1.8	85.9	88.3	1.3	87.1	89.2	2.8	86.7	91.1	3.0	88.4
1-5	87.9	2.9	85.3	90.0	3.3	87.0	89.6	4.3	85.8	90.9	4.7	86.6
1-6	88.8	2.6	86.5	85.2	3.9	81.8	91.1	4.5	87.0	93.6	4.8	89.2
1-7	85.2	4.3	81.5	73.1	4.2	70.0	83.0	4.4	79.4	73.7	5.0	70.0
1-8	88.5	7.4	82.0	92.7	8.6	84.7	90.1	7.9	83.0	93.2	9.9	84.0
1-9	84.7	8.1	77.8	88.4	8.8	80.6	89.7	11.4	79.4	90.0	13.4	77.9
2-1	84.0	3.2	81.3	88.2	3.9	84.8	81.9	3.1	79.4	88.2	4.8	84.0
2-2	70.5	5.5	66.6	82.5	4.9	78.5	74.4	3.9	71.5	82.7	4.1	79.3
2-3	78.3	5.2	74.3	80.4	4.2	77.0	79.7	4.9	75.8	84.6	5.4	80.0
2-4	80.8	5.9	76.0	88.8	7.6	82.0	83.5	4.8	79.5	89.1	6.1	83.8
3-1	86.1	7.1	80.0	88.7	5.9	83.5	86.6	6.3	81.2	88.4	8.3	81.1
3-2	83.4	7.2	77.4	85.7	7.6	79.2	88.2	8.6	80.6	88.3	8.7	80.6
3-3	78.2	10.7	69.8	80.7	10.5	72.2	84.2	14.5	72.0	88.3	12.7	77.1
Mean	84.0	4.9	79.9	86.0	5.1	81.6	86.2	5.7	81.3	88.3	6.3	82.7
MAD	3.6	2.2	4.7	3.7	2.3	4.1	3.8	2.6	4.2	3.2	2.7	4.2

Table 5.6: Performance of Adaptive Certainty Classifier using the real EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_1 **- High Certainty**			e_2 **- Sequential Assignment**			e_3 **- High Certainty**			e_4 **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	84.5	1.4	83.4	74.6	0.9	73.9	83.5	2.7	81.3	74.7	2.1	73.1
2	85.2	9.6	77.0	83.4	19.2	67.4	82.4	18.5	67.2	82.8	24.4	62.7
3	80.7	0.5	80.3	82.1	0.6	81.6	81.0	1.5	79.8	82.3	2.0	80.7
4	88.9	0.4	88.6	87.0	0.0	87.0	87.3	0.5	86.8	90.3	0.7	89.7
5	80.1	2.6	78.0	75.5	2.6	73.5	78.9	6.3	73.9	72.9	8.4	66.8
Mean	83.9	2.9	81.4	80.5	4.7	76.7	82.6	5.9	77.8	80.6	7.5	74.6
MAD	2.8	2.7	3.6	4.4	5.8	6.1	2.2	5.2	5.8	5.5	7.1	8.5

Table 5.7: Performance of Adaptive Fuzzy k-NN Classifier using the independent simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_5 **- High Certainty**			e_6 **- Sequential Assignment**			e_7 **- High Certainty**			e_8 **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	98.7	1.0	97.6	99.7	0.3	99.3	96.6	0.7	96.0	99.3	1.0	98.3
2	98.5	0.3	98.3	99.4	0.6	98.8	97.4	0.9	96.5	99.4	0.6	98.8
3	95.1	1.5	93.7	99.3	0.7	98.5	92.7	2.9	90.0	98.8	1.2	97.6
4	95.1	1.7	93.5	98.7	1.4	97.3	94.8	1.9	93.0	98.0	1.4	96.6
5	97.7	1.6	96.2	99.6	0.4	99.2	97.1	1.4	95.8	98.7	1.0	97.7
6	95.4	1.3	94.2	97.9	1.8	96.1	96.0	1.5	94.6	96.1	1.8	94.4
7	94.9	3.9	91.3	97.0	3.1	94.1	94.6	3.0	91.8	96.9	2.5	94.4
8	95.0	1.8	93.3	98.3	1.4	96.9	94.8	1.8	93.1	98.0	1.6	96.4
9	94.0	3.2	91.1	97.5	1.7	95.8	91.7	4.4	87.6	96.1	2.2	94.0
10	94.1	2.7	91.5	96.4	2.8	93.6	91.5	4.1	87.7	96.1	2.7	93.5
11	93.3	4.7	88.9	95.5	4.2	91.5	92.0	5.6	86.9	95.0	4.1	91.1
12	93.8	3.2	90.8	96.5	2.9	93.7	91.8	5.7	86.6	95.1	4.9	90.5
13	85.2	9.4	77.2	91.8	5.5	86.8	85.5	7.9	78.7	91.5	6.0	86.1
14	93.9	5.1	89.1	97.0	3.6	93.5	93.6	4.8	89.1	96.3	3.9	92.6
15	93.6	4.2	89.6	97.4	3.1	94.4	92.8	5.6	87.6	95.8	3.5	92.5
16	90.3	3.9	86.8	95.5	2.9	92.7	88.6	5.7	83.5	94.7	4.8	90.1
17	92.7	4.9	88.2	96.4	3.5	93.0	91.2	5.2	86.5	96.0	3.1	93.0
18	84.8	11.9	74.5	91.8	10.2	82.4	84.0	13.4	72.7	90.7	10.8	80.9
19	86.6	9.3	78.6	92.1	7.5	85.2	84.6	11.3	75.0	90.7	9.7	81.9
20	88.6	7.3	82.2	94.4	5.9	88.8	88.6	6.2	83.1	93.7	4.8	89.2
Mean	93.1	4.1	89.3	96.6	3.2	93.6	92.0	4.7	87.8	95.8	3.6	92.5
MAD	3.0	2.4	4.9	1.9	1.8	3.5	3.0	2.4	4.8	1.9	2.0	3.7

Table 5.8: Performance of Adaptive Fuzzy k-NN Classifier using the groups of related simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_5 **- High Certainty**			e_6 **- Sequential Assignment**			e_7 **- High Certainty**			e_8 **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	91.4	4.3	87.5	94.5	2.3	92.3	92.1	4.5	87.9	92.2	3.1	89.3
1-2	89.2	4.4	85.3	96.3	3.8	92.7	90.2	4.8	85.9	92.9	5.1	88.2
1-3	89.4	6.6	83.6	88.6	4.5	84.6	86.7	7.3	80.3	90.7	4.9	86.2
1-4	92.0	4.0	88.4	95.6	2.6	93.0	87.2	4.1	83.7	92.4	3.3	89.3
1-5	91.3	8.5	83.6	90.9	5.6	85.8	87.8	6.7	81.9	90.9	5.7	85.7
1-6	93.4	6.7	87.1	94.2	5.7	88.8	91.6	7.5	84.7	93.4	5.9	87.9
1-7	82.2	7.9	75.7	80.5	5.3	76.3	87.1	9.7	78.6	76.6	7.4	70.9
1-8	92.3	9.6	83.5	97.4	6.7	90.8	92.4	9.9	83.2	90.5	6.3	84.7
1-9	90.9	11.4	80.6	76.1	9.6	68.8	90.4	11.8	79.7	86.8	11.1	77.1
2-1	92.0	6.9	85.7	94.9	3.8	91.3	91.3	7.1	84.8	96.3	4.8	91.7
2-2	83.2	6.1	78.1	92.0	5.3	87.1	82.0	8.5	75.1	90.0	4.9	85.6
2-3	86.0	8.5	78.7	90.7	3.5	87.5	88.2	10.6	78.8	88.2	3.6	84.9
2-4	91.7	11.0	81.7	89.9	3.9	86.4	94.2	9.6	85.1	97.9	6.3	91.7
3-1	87.8	8.5	80.3	93.6	7.3	86.8	92.0	9.6	83.2	92.4	7.6	85.3
3-2	91.9	11.5	81.3	94.0	8.7	85.9	93.0	11.2	82.6	87.7	8.0	80.6
3-3	87.8	18.0	72.0	93.8	11.0	83.4	91.4	16.6	76.2	95.0	11.6	84.0
Mean	89.5	8.4	82.1	91.4	5.6	86.3	89.8	8.7	82.0	90.8	6.2	85.2
MAD	2.6	2.5	3.5	4.0	1.9	4.2	2.6	2.3	2.8	3.3	1.8	3.6

Table 5.9: Performance of Adaptive Fuzzy k-NN Classifier using the real EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_5 **- High Certainty**			e_6 **- Sequential Assignment**			e_7 **- High Certainty**			e_8 **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	92.0	4.4	87.9	76.4	3.3	73.9	91.0	5.5	86.0	77.5	4.9	73.7
2	90.7	10.4	81.3	94.7	20.7	75.1	90.3	19.3	72.9	91.4	25.9	67.8
3	94.1	4.3	90.0	94.5	4.4	90.3	91.4	5.8	86.1	93.7	5.4	88.6
4	96.5	3.9	92.7	96.7	0.8	95.9	94.3	3.7	90.8	98.1	2.1	96.0
5	92.9	8.8	84.7	93.7	6.5	87.6	90.2	12.7	78.7	90.4	5.8	85.2
Mean	93.2	6.4	87.3	91.2	7.1	84.6	91.4	9.4	82.9	90.2	8.8	82.2
MAD	1.6	2.6	3.5	5.9	5.4	8.0	1.1	5.3	5.7	5.1	6.8	9.2

Table 5.10: Performance of Adaptive Normalized Cross-Correlation Classifier using the independent simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_9 **- High Certainty**			e_{10} **- Sequential Assignment**			e_{11} **- High Certainty**			e_{12} **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	99.7	3.7	96.0	99.7	2.7	97.0	99.3	2.0	97.3	99.7	2.4	97.3
2	98.8	1.5	97.4	98.8	1.5	97.4	98.3	0.9	97.4	98.9	1.2	97.7
3	95.9	4.1	92.0	97.8	3.7	94.2	93.9	5.2	89.1	97.3	4.0	93.4
4	98.4	6.8	91.7	97.1	6.2	91.0	91.7	6.1	86.1	96.9	5.3	91.7
5	98.7	1.2	97.5	99.4	1.2	98.3	98.8	1.8	97.1	99.2	1.6	97.7
6	96.5	3.8	92.8	96.5	3.3	93.3	96.7	4.4	92.5	94.9	3.5	91.6
7	93.0	4.1	89.2	96.2	4.2	92.1	93.9	5.0	89.2	96.3	5.1	91.4
8	96.4	4.7	91.9	96.1	3.4	92.8	96.6	3.7	93.0	97.3	2.9	94.5
9	96.3	4.2	92.3	97.6	2.1	95.5	90.0	9.4	81.5	91.7	7.8	84.5
10	96.1	4.5	91.8	96.4	3.2	93.2	95.8	5.1	90.9	96.2	4.2	92.2
11	91.9	7.2	85.3	83.6	8.0	76.9	91.6	9.7	82.7	93.2	11.5	82.5
12	93.8	6.6	87.7	96.5	7.3	89.4	93.2	9.0	84.8	91.6	8.0	84.3
13	91.5	11.3	81.2	93.2	7.3	86.4	91.9	9.7	83.0	92.3	7.6	85.2
14	95.6	7.6	88.3	96.0	7.9	88.4	93.4	9.3	84.7	95.4	7.5	88.2
15	92.2	9.3	83.6	93.1	9.7	84.1	91.1	8.6	83.3	86.6	9.5	78.3
16	87.8	10.8	78.3	89.8	9.7	81.1	83.7	10.5	75.0	91.0	14.6	77.7
17	92.8	7.7	85.7	92.7	8.7	84.6	92.2	9.7	83.2	92.7	12.3	81.3
18	83.2	21.7	65.1	82.5	21.6	64.7	75.6	25.7	56.2	83.0	25.7	61.6
19	84.5	14.4	72.3	90.8	16.5	75.8	89.4	19.6	71.8	90.1	16.3	75.4
20	84.2	12.6	73.6	90.3	13.8	77.9	91.3	16.9	75.9	89.4	16.1	75.0
Mean	93.4	7.4	86.7	94.2	7.1	87.7	92.4	8.6	84.7	93.7	8.4	86.1
MAD	3.9	3.6	6.8	3.8	3.9	7.0	3.6	4.3	7.0	3.5	4.8	7.5

Table 5.11: Performance of Adaptive Normalized Cross-Correlation Classifier using the groups of related simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_9 **- High Certainty**			e_{10} **- Sequential Assignment**			e_{11} **- High Certainty**			e_{12} **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	94.9	7.5	87.8	96.2	6.3	90.1	95.4	8.3	87.5	93.9	7.8	86.6
1-2	86.2	6.4	80.7	91.5	9.7	82.7	86.6	8.8	79.0	86.7	9.3	78.6
1-3	92.2	11.7	81.4	89.4	9.9	80.6	88.9	9.6	80.3	83.7	9.1	76.1
1-4	92.8	6.1	87.1	89.4	4.8	85.1	88.4	6.2	82.9	90.8	6.3	85.1
1-5	78.5	7.2	72.9	81.9	19.5	65.9	79.2	11.7	70.0	78.2	21.0	61.8
1-6	92.5	17.5	76.3	80.4	17.2	66.6	95.4	19.0	77.3	86.5	17.7	71.2
1-7	89.4	19.5	72.0	82.9	25.9	61.4	76.4	14.3	65.5	83.3	23.9	63.3
1-8	96.6	23.1	74.3	96.6	22.1	75.2	82.8	17.2	68.5	95.2	27.4	69.1
1-9	82.6	17.4	68.2	87.9	27.3	63.9	91.9	23.1	70.6	76.5	21.5	60.0
2-1	89.8	8.3	82.3	91.5	7.1	85.0	93.5	8.7	85.3	95.2	6.3	89.3
2-2	81.4	10.4	72.9	84.6	9.4	76.6	93.6	20.7	74.2	86.6	11.4	76.8
2-3	90.4	20.2	72.1	91.0	15.8	76.6	92.9	11.9	81.6	90.2	14.6	77.0
2-4	88.6	13.1	77.0	82.1	17.9	67.4	92.2	14.0	79.3	77.7	10.0	69.9
3-1	91.8	21.3	72.3	78.0	24.4	59.0	87.5	23.2	67.2	90.7	23.1	69.7
3-2	86.6	27.3	63.0	82.8	34.6	54.2	85.6	21.4	67.3	75.0	22.8	57.9
3-3	79.4	30.6	55.1	73.1	20.5	57.4	88.6	29.8	62.2	93.4	29.5	65.9
Mean	88.4	15.5	74.7	86.2	17.1	71.7	88.7	15.5	74.9	86.5	16.4	72.4
MAD	4.4	6.6	6.2	5.5	7.1	9.8	4.3	5.7	6.7	5.6	7.0	7.8

Table 5.12: Performance of Adaptive Normalized Cross-Correlation Classifier using the real EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_9 **- High Certainty**			e_{10} **- Sequential Assignment**			e_{11} **- High Certainty**			e_{12} **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	92.1	9.5	83.4	78.0	7.1	72.5	90.3	13.6	78.0	85.3	11.2	75.8
2	89.5	18.4	73.0	82.6	20.6	65.6	87.4	23.7	66.7	88.0	27.7	63.6
3	93.3	6.2	87.4	94.5	6.6	88.3	92.0	10.0	82.8	94.4	8.7	86.2
4	82.2	3.5	79.4	92.7	17.6	76.3	97.5	13.2	84.6	93.8	18.6	76.3
5	93.7	17.7	77.1	92.9	14.9	79.1	93.3	26.5	68.5	90.5	41.2	53.2
Mean	90.2	11.1	80.1	88.1	13.4	76.4	92.1	17.4	76.1	90.4	21.5	71.0
MAD	3.4	5.6	4.3	6.3	5.2	5.9	2.6	6.2	6.8	3.0	10.4	10.1

Table 5.13: Performance of Adaptive Pseudo Correlation Classifier using the independent simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_{13} **- High Certainty**			e_{14} **- Sequential Assignment**			e_{15} **- High Certainty**			e_{16} **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	93.9	0.7	93.3	92.3	1.1	91.2	98.0	2.1	96.0	98.7	1.7	97.0
2	96.2	0.9	95.3	96.5	0.6	95.9	95.6	0.6	95.1	96.5	0.6	95.9
3	80.8	1.2	79.8	88.6	1.1	87.6	90.5	2.4	88.3	93.7	1.8	92.0
4	88.1	2.5	85.9	90.1	1.5	88.8	95.3	4.7	90.8	94.6	4.0	90.8
5	96.5	2.0	94.6	97.3	1.2	96.2	96.5	1.0	95.6	97.5	1.2	96.3
6	91.8	2.5	89.5	93.3	2.4	91.1	94.2	3.0	91.4	93.3	2.4	91.1
7	84.6	2.7	82.3	84.3	2.7	82.0	90.0	3.1	87.2	91.4	3.3	88.5
8	93.8	2.5	91.4	94.1	1.7	92.5	93.6	3.5	90.3	94.1	4.2	90.2
9	87.6	3.1	85.0	89.6	6.0	84.2	90.0	6.3	84.4	91.2	7.8	84.1
10	89.2	4.1	85.5	90.8	5.5	85.8	88.6	8.2	81.4	91.4	8.0	84.1
11	88.5	5.2	83.9	89.2	6.2	83.6	91.6	6.0	86.1	92.7	12.0	81.6
12	88.2	7.8	81.3	88.2	7.8	81.3	91.6	9.7	82.7	88.6	9.8	79.9
13	79.7	9.0	72.6	80.6	9.8	72.7	84.9	10.3	76.2	85.2	18.6	69.4
14	90.4	7.7	83.5	90.7	6.6	84.7	93.0	7.0	86.5	93.6	5.9	88.1
15	83.4	4.6	79.5	85.8	4.5	81.9	88.4	8.8	80.6	88.7	6.8	82.6
16	86.0	7.5	79.5	86.1	14.4	73.8	86.4	9.2	78.4	88.0	6.0	82.7
17	85.1	4.5	81.3	85.1	3.7	82.0	87.9	6.4	82.3	90.0	6.5	84.2
18	76.9	13.2	66.7	78.0	16.5	65.1	76.8	19.9	61.5	83.1	25.7	61.7
19	74.2	7.6	68.6	80.4	11.4	71.3	82.7	13.5	71.5	85.6	13.8	73.8
20	81.2	6.4	76.0	82.4	7.5	76.2	88.1	9.6	79.7	89.5	6.3	83.9
Mean	86.8	4.8	82.8	88.2	5.6	83.4	90.2	6.8	84.3	91.4	7.3	84.9
MAD	4.9	2.6	6.0	4.3	3.6	6.4	3.8	3.5	6.4	3.2	4.4	6.6

Table 5.14: Performance of Adaptive Pseudo Correlation Classifier using the groups of related simulated EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_{13} **- High Certainty**			e_{14} **- Sequential Assignment**			e_{15} **- High Certainty**			e_{16} **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	89.4	5.1	84.8	88.7	3.2	85.9	92.3	8.7	84.3	90.9	7.5	84.1
1-2	86.3	5.1	82.0	89.3	10.3	80.1	87.6	7.3	81.2	91.7	8.8	83.7
1-3	86.3	12.9	75.2	86.8	15.1	73.7	86.0	8.0	79.1	84.8	7.6	78.4
1-4	87.2	6.9	81.2	85.1	8.8	77.5	90.2	7.0	83.9	86.9	6.0	81.6
1-5	87.4	9.0	79.6	86.3	10.0	77.6	87.5	8.8	79.8	87.0	13.3	75.4
1-6	88.7	8.0	81.6	76.8	12.5	67.2	87.2	9.4	79.0	85.5	15.8	72.0
1-7	75.2	11.5	66.5	69.2	8.9	63.0	86.8	13.5	75.1	81.6	20.1	65.2
1-8	89.2	18.8	72.4	88.1	22.7	68.1	87.0	14.4	74.5	83.6	15.8	70.4
1-9	92.0	21.6	72.1	81.5	28.2	58.5	89.5	19.7	71.9	77.6	18.7	63.1
2-1	89.6	13.7	77.3	85.7	8.0	78.9	75.3	6.3	70.5	84.2	6.9	78.4
2-2	83.4	8.7	76.1	84.4	8.8	76.9	73.2	6.7	68.3	80.5	20.6	63.9
2-3	84.4	9.0	76.8	88.7	9.0	80.7	88.5	10.3	79.4	83.9	7.9	77.3
2-4	81.8	9.1	74.4	76.2	8.2	69.9	84.0	10.8	74.9	89.4	13.1	77.7
3-1	87.6	12.6	76.2	87.3	13.1	75.9	81.2	17.9	66.6	84.5	19.4	68.2
3-2	83.4	18.7	67.8	87.1	25.1	65.2	90.3	23.1	69.4	91.1	22.8	70.4
3-3	69.7	25.1	51.9	81.5	20.8	64.5	82.3	28.7	58.6	81.9	24.3	62.0
Mean	85.1	12.2	74.8	83.9	13.3	72.7	86.5	12.6	74.8	85.4	14.3	73.2
MAD	4.1	4.7	5.4	4.3	5.7	6.6	4.0	5.3	5.4	3.2	5.4	6.3

Table 5.15: Performance of Adaptive Pseudo Correlation Classifier using the real EMG signals with two types of features and two types of classifier seeding.

	First-order discrete derivative features						**WT of first-order discrete derivative features**					
	e_{13} **- High Certainty**			e_{14} **- Sequential Assignment**			e_{15} **- High Certainty**			e_{16} **- Sequential Assignment**		
EMG signal	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	85.6	4.3	81.9	84.8	3.6	81.8	87.5	5.1	83.1	88.7	4.6	84.7
2	88.5	23.0	68.2	82.2	26.3	60.6	79.9	23.4	61.3	76.8	31.9	52.3
3	83.7	2.7	81.4	83.4	3.0	80.9	83.3	6.2	78.1	84.0	7.0	78.1
4	91.9	1.4	90.6	92.4	1.5	91.0	92.4	1.9	90.6	94.3	1.9	92.5
5	80.5	8.0	74.0	73.3	13.1	63.7	84.0	11.1	74.6	71.8	29.6	50.5
Mean	86.0	7.9	79.2	83.2	9.5	75.6	85.4	9.5	77.5	83.1	15.0	71.6
MAD	3.3	6.1	6.5	4.4	8.1	10.8	3.6	6.2	7.7	7.1	12.6	16.2

5.3.2 Sensitivity Analysis

We performed a sensitivity analysis for deciding how many base classifiers to use in an ensemble.

An exhaustive search for the best performing classifier ensemble was performed and we concluded that when choosing an ensemble of six base classifiers gives better performing ensemble as we tried different base classifiers size ensembles.

5.3.3 Fixed Classifier Ensemble Results

An ensemble of six base classifiers $e_1, e_2, e_5, e_6, e_7, e_8$ was used for carrying on the analysis. These classifiers were chosen to work as a team as they demonstrate the best classification accuracy. Two base e_1, e_2 classifiers were adaptive certainty-based classifiers (ACC) [85] and four classifiers e_5, e_6, e_7, e_8 were adaptive fuzzy k-NN classifiers (AFNNC) [83]. Both ACC classifiers e_1, e_2 were fed with time-domain first-order discrete derivative features. Two of the AFNNC classifiers e_5, e_6 were fed with time-domain first-order discrete derivative features, while the other two e_7, e_8 were fed with wavelet-domain first-order discrete derivative features. ACC classifier e_1 and AFNNC Classifiers e_5, e_7 were seeded with MUPs having the highest shape certainty [85]. ACC classifier e_2 and AFNNC classifiers e_6, e_8 were seeded with a number of isolated MUPs sequentially selected for each MUPT.

The performance with respect to each signal, the mean performance and the mean absolute deviation (MAD) of the performance are reported. Results of the one level classifier fusion scheme, namely the majority voting, the average fixed combination rule, and the Sugeno fuzzy integral across the used EMG data set are presented in Tables 5.16, 5.17, and 5.18. Results of the hybrid classifier fusion scheme, i.e., the hybrid combiner consisting of majority voting and average fixed combination rule and the hybrid combiner consisting of majority voting and Sugeno fuzzy integral are also presented in Tables 5.16, 5.17, and 5.18, respectively.

Table 5.16: Performance of fixed ensemble classifier fusion system using the independent simulated EMG signals.

EMG signal	**Single Combiner Classifier Fusion**									**Hybrid Combiners Classifier Fusion**					
	Adaptive Majority Voting			**Adaptive Average Fixed Rule**			**Adaptive Sugeno Fuzzy Integral**			**Adaptive Majority Voting with Average Fixed Rule**			**Adaptive Majority Voting with Sugeno Fuzzy Integral**		
	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	97.3	0.0	97.3	98.7	0.3	98.3	98.3	0.7	97.6	99.0	0.3	98.7	99.0	0.3	98.7
2	98.0	0.3	97.7	98.5	0.6	98.0	97.4	1.2	96.2	98.5	0.6	98.0	98.0	0.9	97.1
3	95.6	0.8	94.9	97.6	0.7	96.8	85.4	1.1	84.4	98.1	1.0	97.1	97.6	1.0	96.6
4	94.6	0.5	94.2	97.1	1.6	95.5	95.5	2.8	92.8	97.5	1.4	96.2	96.0	2.1	93.9
5	96.7	0.4	96.3	98.3	1.0	97.3	97.3	1.6	95.8	98.7	0.8	97.9	97.7	1.2	96.5
6	95.6	0.7	94.9	95.8	1.6	94.2	94.0	3.7	90.5	97.2	1.1	96.1	96.3	1.8	94.6
7	93.0	1.9	91.3	95.6	3.3	92.5	95.5	3.8	91.8	96.2	2.2	94.1	96.2	2.4	93.9
8	94.4	0.7	93.8	96.4	1.5	95.0	95.0	3.0	92.2	96.6	1.3	95.3	95.9	2.0	94.1
9	92.7	0.6	92.1	94.9	2.4	92.7	92.6	3.9	89.0	96.0	1.4	94.6	95.1	2.2	93.0
10	92.2	0.4	91.8	94.1	2.4	91.8	91.9	4.4	87.8	95.4	1.4	94.1	94.3	2.3	92.2
11	90.7	2.0	88.9	93.4	4.8	88.9	87.6	5.0	83.2	90.1	3.0	87.4	93.3	4.1	89.4
12	91.1	1.5	89.7	93.3	3.4	90.2	92.2	4.8	87.8	94.6	2.7	92.1	93.5	3.4	90.3
13	82.7	3.3	80.0	89.0	7.0	82.8	86.1	9.4	77.9	90.5	4.6	86.3	88.2	5.9	83.0
14	92.0	2.4	89.8	94.3	3.5	91.0	92.6	5.0	88.0	95.4	3.1	92.5	94.8	4.0	91.0
15	91.1	2.0	89.3	94.1	3.7	90.7	92.6	5.2	87.8	95.2	3.4	91.9	94.1	3.9	90.5
16	87.2	0.9	86.4	92.0	4.0	88.3	88.5	6.0	83.2	92.9	2.8	90.3	91.0	4.2	87.2
17	90.0	2.3	87.9	93.1	4.3	89.1	91.4	6.0	86.0	94.4	3.1	91.5	93.0	4.6	88.8
18	78.3	4.8	74.5	87.1	8.4	79.8	85.8	10.6	76.7	88.3	7.1	82.0	87.6	8.4	80.2
19	77.7	2.4	75.9	86.9	6.4	81.3	85.7	8.4	78.5	88.3	5.1	83.7	86.9	5.3	82.3
20	86.2	2.3	84.3	89.9	5.7	84.7	88.2	7.4	81.6	91.8	4.3	87.9	90.8	5.4	85.9
Mean	90.9	1.5	89.5	94.0	3.3	90.9	91.7	4.7	87.4	94.7	2.5	92.4	94.0	3.3	91.0
MAD	4.3	1.0	4.9	2.7	1.8	4.3	3.5	2.1	4.8	2.7	1.4	3.8	2.7	1.7	4.1

Table 5.17: Performance of fixed ensemble classifier fusion system using the groups of related simulated EMG signals.

EMG signal	**Single Combiner Classifier Fusion**									**Hybrid Combiners Classifier Fusion**					
	Adaptive Majority Voting			**Adaptive Average Fixed Rule**			**Adaptive Sugeno Fuzzy Integral**			**Adaptive Majority Voting with Average Fixed Rule**			**Adaptive Majority Voting with Sugeno Fuzzy Integral**		
	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	90.9	1.0	90.0	92.3	2.9	89.7	91.6	4.5	87.5	93.3	2.0	91.5	92.5	2.5	90.2
1-2	88.9	1.6	87.5	91.7	3.0	89.0	89.8	5.9	84.5	92.1	2.4	89.9	91.6	4.0	87.9
1-3	85.5	2.4	83.4	90.0	4.0	86.4	88.6	6.2	83.1	91.0	3.2	88.2	91.3	4.5	87.1
1-4	90.0	1.4	88.7	92.2	3.3	89.2	89.3	5.4	84.5	93.7	2.3	91.6	92.4	3.1	89.5
1-5	87.2	2.8	84.8	93.0	5.1	88.2	91.1	4.9	86.6	92.6	4.4	88.6	92.3	4.1	88.5
1-6	87.6	2.5	85.4	96.1	4.9	91.4	96.2	5.7	90.8	95.2	4.1	91.3	95.7	5.0	90.9
1-7	76.3	2.5	74.4	87.3	5.3	82.7	82.3	6.1	77.3	86.8	5.1	82.4	81.1	5.6	76.6
1-8	87.1	3.6	84.0	94.8	5.9	89.2	94.3	6.0	88.6	95.1	5.8	89.6	94.3	6.1	88.5
1-9	82.7	4.6	78.9	92.2	8.3	84.6	91.5	9.1	83.2	93.1	7.7	85.9	94.0	8.5	85.9
2-1	88.1	2.3	86.0	93.7	3.8	90.1	92.7	5.0	88.1	93.9	3.1	91.0	92.8	4.0	89.1
2-2	81.0	3.3	78.3	85.8	5.3	81.2	83.6	6.5	78.1	88.0	4.4	84.1	87.5	5.8	82.4
2-3	79.0	1.7	77.7	89.7	5.3	84.9	86.8	6.2	81.4	88.7	3.4	85.6	87.5	4.4	83.6
2-4	86.9	3.4	84.0	94.4	4.9	89.8	92.4	5.0	87.8	94.7	4.9	90.1	94.7	5.6	89.4
3-1	86.2	4.4	82.4	93.3	6.4	87.3	93.4	7.2	86.6	93.3	6.2	87.5	93.2	6.2	87.4
3-2	84.6	4.5	80.8	95.5	7.7	88.2	95.0	7.7	87.7	95.3	7.5	88.2	94.6	7.5	87.5
3-3	77.9	5.9	73.3	92.7	9.6	83.8	90.6	10.0	81.6	92.8	9.7	83.8	92.9	10.5	83.2
Mean	85.0	3.0	82.5	92.2	5.4	87.2	90.6	6.3	84.8	92.5	4.8	88.1	91.8	5.5	86.7
MAD	3.6	1.1	3.9	2.0	1.4	2.5	2.9	1.1	3.1	2.0	1.7	2.4	2.5	1.5	2.7

Table 5.18: Performance of fixed ensemble classifier fusion system using the real EMG signals.

EMG signal	**Single Combiner Classifier Fusion**									**Hybrid Combiners Classifier Fusion**					
	Adaptive Majority Voting			**Adaptive Average Fixed Rule**			**Adaptive Sugeno Fuzzy Integral**			**Adaptive Majority Voting with Average Fixed Rule**			**Adaptive Majority Voting with Sugeno Fuzzy Integral**		
	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	77.3	1.9	75.9	92.9	3.5	89.6	91.6	4.1	87.8	92.7	3.1	89.8	91.9	3.9	88.3
2	71.5	3.5	69.0	90.3	8.7	82.4	87.4	11.6	77.3	91.8	9.0	83.5	89.8	10.5	80.4
3	86.2	1.0	85.4	92.5	2.4	90.3	91.8	4.8	87.4	92.5	2.4	90.3	92.2	2.9	89.5
4	92.5	0.3	92.2	97.9	1.0	97.0	97.5	2.8	94.8	97.9	1.0	97.0	97.3	2.1	95.2
5	84.5	2.7	82.2	94.5	5.1	89.7	92.4	6.3	86.6	93.8	4.1	89.9	92.8	4.4	88.7
Mean	82.4	1.9	80.9	93.6	4.1	89.8	92.1	5.9	86.8	93.7	3.9	90.1	92.8	4.8	88.4
MAD	6.4	1.0	6.8	2.1	2.2	3.1	2.2	2.4	3.9	1.7	2.1	2.8	1.8	2.3	3.3

5.3.4 Diversity-Based Classifier Selection Ensemble Results

With the diversity-based hybrid classifier fusion scheme, we performed two experiments.

In the first experiment, we used a base classifier pool containing eight base classifiers $e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8$ from which we selected six classifiers to work as a team in the ensemble for every signal at each stage combiners. The number of classifier ensembles that can be created is $\binom{8}{6} = 28$ ensembles.

Four base e_1, e_2, e_3, e_4 classifiers were adaptive certainty-based classifiers (ACC) [85] and the other four classifiers e_5, e_6, e_7, e_8 were adaptive fuzzy k-NN classifiers (AFNNC) [83]. Two of the ACC classifiers e_1, e_2 were fed with time-domain first-order discrete derivative features, while the other two e_3, e_4 were fed with wavelet-domain first-order discrete derivative features. Two of the AFNNC classifiers e_5, e_6 were fed with time-domain first-order discrete derivative features, while the other two e_7, e_8 were fed with wavelet-domain first-order discrete derivative features. ACC Classifiers e_1, e_3 and AFNNC Classifiers e_5, e_7 were seeded with MUPs having the highest shape certainty [85]. ACC Classifiers e_2, e_4 and AFNNC Classifiers e_6, e_8 were seeded with a number of isolated MUPs sequentially selected for each MUPTs.

The performance with respect to each signal, the mean performance and the mean absolute deviation (MAD) of the performance are reported. Results of the diversity-based hybrid classifier fusion scheme, i.e., the hybrid combiner consisting of majority voting and average fixed combination rule and the hybrid combiner consisting of majority voting and Sugeno fuzzy integral with the pre-stage classifier selection module for each combiner are also presented in Tables 5.19, 5.20, and 5.21.

In the second experiment, we used a base classifier pool containing sixteen base classifiers $e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11}, e_{12}, e_{13}, e_{14}, e_{15}, e_{16}$ from which we selected six classifiers to work as a team in the ensemble for every signal at each stage combiners. The number of classifier ensembles that can be created is $\binom{16}{6} = 8008$ ensembles. The used base classifiers are the same as the classifiers described in Section 5.3.1.

Table 5.19: Performance of 6 out of 8 diversity-based ensemble classifier fusion system using the independent simulated EMG signals.

EMG signal	Ensemble Classifier IDs	Max Kappa Value	First Combiner Majority Voting			Ensemble Classifier IDs	Unassigned Min Kappa	**Adaptive MV with Average Fixed Rule**			**Adaptive MV with Sugeno Fuzzy Integral**		
			$A_r\%$	$E_r\%$	$CC_r\%$			$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	3 4 5 6 7 8	0.94	97.3	0.0	97.3	1 4 5 6 7 8	-0.13	98.7	0.3	98.3	98.7	0.7	98.0
2	1 3 5 6 7 8	0.97	98.3	0.3	98.0	1 2 3 5 6 7	-0.18	98.8	0.6	98.3	98.3	0.9	97.4
3	1 3 5 6 7 8	0.90	96.6	0.8	95.9	1 3 4 5 6 7	-0.14	97.8	0.7	97.1	97.8	1.0	96.8
4	3 4 5 6 7 8	0.89	94.8	0.7	94.2	2 4 5 6 7 8	-0.03	97.8	1.4	96.4	97.1	2.1	95.1
5	1 2 3 4 5 8	0.96	95.2	0.2	95.0	1 2 4 5 7 8	-0.15	98.3	0.4	97.9	98.1	0.8	97.3
6	1 2 3 4 6 8	0.94	92.8	0.2	92.6	2 3 4 5 6 8	-0.13	96.0	0.5	95.4	95.6	1.3	94.4
7	3 4 5 6 7 8	0.89	93.4	2.1	91.4	2 3 4 5 7 8	-0.14	96.9	2.3	94.6	97.0	2.2	94.9
8	1 2 3 4 6 8	0.94	93.0	0.5	92.5	1 2 4 5 6 7	-0.17	95.9	1.1	94.8	95.8	1.3	94.5
9	1 2 3 4 5 7	0.89	90.6	0.3	90.3	1 2 3 5 6 7	-0.14	96.1	1.2	94.9	95.8	1.2	94.6
10	1 2 3 4 5 7	0.91	91.8	0.6	91.2	1 3 4 5 6 7	-0.14	95.0	1.0	94.1	94.9	1.4	93.5
11	2 4 5 6 7 8	0.88	90.2	2.0	88.4	2 3 4 5 6 7	-0.08	94.8	2.9	92.1	94.5	3.8	90.9
12	1 2 3 4 7 8	0.88	89.6	1.3	88.4	1 2 3 6 7 8	-0.12	94.3	1.9	92.6	93.6	2.4	91.3
13	1 2 3 4 5 7	0.82	82.0	1.2	81.0	1 2 3 4 5 7	-0.14	90.6	3.3	87.6	90.3	3.4	87.2
14	1 3 5 6 7 8	0.90	93.1	2.4	90.9	1 3 4 5 7 8	0.06	95.5	3.2	92.5	94.8	3.6	91.5
15	3 4 5 6 7 8	0.86	91.4	2.4	89.2	2 3 4 5 6 7	-0.06	95.2	2.9	92.5	94.6	3.1	91.6
16	1 2 3 4 5 7	0.85	86.2	0.8	85.5	1 2 3 5 6 7	-0.12	93.1	2.4	90.9	92.3	2.6	89.9
17	3 4 5 6 7 8	0.86	90.5	2.4	88.3	2 3 4 5 6 7	-0.09	95.0	3.2	92.0	94.9	3.6	91.5
18	1 2 3 4 6 8	0.74	70.5	4.1	67.5	2 3 5 6 7 8	-0.06	85.5	6.3	80.2	84.9	7.4	78.6
19	1 2 3 4 5 7	0.45	78.3	2.0	76.7	1 2 4 5 7 8	-0.11	86.7	4.6	82.7	86.0	5.4	83.5
20	1 2 3 4 5 7	0.86	85.1	1.1	84.1	1 2 3 5 7 8	-0.13	91.0	3.4	87.9	90.2	3.8	86.8
		Mean	90.0	1.3	88.9			94.7	2.2	92.6	94.3	2.6	92.0
		MAD	4.9	0.9	5.1			2.7	1.3	3.5	2.8	1.3	3.7

Table 5.20: Performance of 6 out of 8 diversity-based ensemble classifier fusion system using the groups of related simulated EMG signals.

EMG signal	Ensemble Classifier IDs	Max Kappa Value	First Combiner Majority Voting			Ensemble Classifier IDs	Unassigned Min Kappa	**Adaptive MV with Average Fixed Rule**			**Adaptive MV with Sugeno Fuzzy Integral**		
			$A_r\%$	$E_r\%$	$CC_r\%$			$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	1 2 3 4 5 7	0.90	91.1	1.0	90.2	1 2 4 5 7 8	-0.08	93.0	2.1	91.0	92.6	2.1	90.7
1-2	1 2 3 4 5 7	0.88	90.7	1.3	89.6	1 2 3 5 7 8	-0.06	92.7	2.1	90.7	92.0	2.7	89.4
1-3	1 2 3 4 5 7	0.84	85.6	1.6	84.3	2 3 4 5 6 7	-0.07	91.4	3.3	88.4	90.7	3.4	87.6
1-4	1 2 3 4 5 7	0.89	90.2	1.3	89.1	1 2 4 5 6 7	-0.12	93.4	2.1	91.4	92.8	2.7	90.3
1-5	1 2 3 4 5 7	0.89	88.8	2.6	86.5	1 2 3 4 5 7	-0.07	91.1	3.7	87.8	90.5	4.2	86.7
1-6	2 3 4 6 7 8	0.87	88.6	2.8	86.1	1 2 4 6 7 8	-0.08	94.6	4.5	90.4	94.9	4.6	90.5
1-7	1 2 3 4 5 8	0.75	74.6	1.9	73.2	1 2 3 4 5 8	-0.09	83.6	3.3	80.9	83.0	4.0	79.7
1-8	1 2 4 5 6 8	0.84	86.7	4.2	83.0	1 2 4 5 6 8	-0.02	95.2	6.5	89.0	95.2	6.5	89.0
1-9	1 2 3 4 6 8	0.81	82.8	6.3	77.6	1 2 4 6 7 8	-0.03	93.0	9.1	84.6	93.1	9.1	84.7
2-1	1 3 5 6 7 8	0.84	89.9	2.3	87.9	1 2 5 6 7 8	-0.02	94.2	3.1	91.3	93.9	3.8	90.3
2-2	1 3 5 6 7 8	0.79	84.7	3.6	81.7	2 3 4 5 6 7	-0.08	88.3	4.2	84.6	87.1	5.1	82.7
2-3	1 2 3 4 5 7	0.80	82.1	1.9	80.5	2 3 4 5 6 7	-0.10	89.3	3.6	86.1	88.7	4.0	85.1
2-4	1 2 3 4 5 7	0.82	86.0	2.3	84.0	1 3 5 6 7 8	-0.06	94.5	4.5	90.2	94.2	3.7	90.7
3-1	1 2 3 4 5 7	0.85	86.5	3.6	83.4	2 3 4 5 6 7	-0.07	92.4	5.5	87.4	92.4	5.8	87.0
3-2	1 2 3 4 6 7	0.81	84.1	4.7	80.2	2 3 4 5 6 8	-0.04	94.1	7.0	87.6	93.8	6.9	87.3
3-3	1 2 3 4 5 7	0.76	81.3	5.1	77.2	1 2 3 5 7 8	-0.03	93.3	9.4	84.5	92.8	9.8	83.7
		Mean	85.9	2.9	83.4			92.1	4.6	87.9	91.7	4.9	87.2
		MAD	3.2	1.3	3.8			2.1	1.8	2.4	2.3	1.7	2.6

Table 5.21: Performance of 6 out of 8 diversity-based ensemble classifier fusion system using the real EMG signals.

EMG signal	Ensemble Classifier IDs	Max Kappa Value	First Combiner Majority Voting			Ensemble Classifier IDs	Unassigned Min Kappa	**Adaptive MV with Average Fixed Rule**			**Adaptive MV with Sugeno Fuzzy Integral**		
			$A_r\%$	$E_r\%$	$CC_r\%$			$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	1 2 3 4 5 7	0.80	75.6	0.9	75.0	1 2 4 5 6 7	-0.16	88.9	2.4	86.8	88.9	2.4	86.8
2	1 2 4 5 6 8	0.66	74.3	3.5	71.6	1 2 5 6 7 8	0.03	92.9	8.2	85.3	91.7	9.5	83.0
3	1 2 3 4 6 8	0.85	82.2	0.4	81.9	1 2 4 6 7 8	-0.13	90.7	2.5	88.4	89.2	2.0	87.4
4	2 3 5 6 7 8	0.88	92.9	0.3	92.5	1 4 5 6 7 8	-0.07	97.8	1.3	96.5	97.3	2.4	94.9
5	1 2 3 4 6 8	0.78	83.3	2.4	81.3	1 2 3 4 6 8	-0.03	94.0	4.8	89.6	92.9	4.8	88.4
		Mean	81.7	1.5	80.5			92.9	3.8	89.3	92.0	4.2	88.1
		MAD	5.4	1.2	5.7			2.5	2.1	3.0	2.5	2.4	2.9

The performance with respect to each signal, the mean performance and the mean absolute deviation (MAD) of the performance are reported. Results of the diversity-based hybrid classifier fusion scheme, i.e., the hybrid combiner consisting of majority voting and average fixed combination rule and the hybrid combiner consisting of majority voting and Sugeno fuzzy integral with the pre-stage classifier selection module for each combiner are also presented in Tables 5.22, 5.23, and 5.24.

5.3.5 Summary of Data Analysis Results

Table 5.25 and Table 5.26 show the mean performance and the mean absolute deviation (MAD) of all the base classifiers and the classifier fusion systems, respectively.

The abbreviations employed in Table 5.26 are defined as:

Best base - The base classifier with the best performance,

Weakest base - The base classifier with the weakest performance,

Mean base - The mean performance of the 16 base classifiers,

AMV - Adaptive majority voting,

AAFR - Adaptive average fixed rule,

ASFI - Adaptive Sugeno fuzzy integral,

AMVAFR - Adaptive majority voting with average fixed rule,

AMVSFI - Adaptive majority voting with Sugeno fuzzy integral,

AMVAFRD - Adaptive diversity-based majority voting with average fixed rule,

AMVSFID - Adaptive diversity-based majority voting with Sugeno fuzzy integral,

SC - Single combiner,

HC - Hybrid combiner,

6/8 - Selecting 6 base classifiers from the classifier pool containing 8 classifiers,

6/16 - Selecting 6 base classifiers from the classifier pool containing 16 classifiers.

Table 5.22: Performance of 6 out of 16 diversity-based ensemble classifier fusion system using the independent simulated EMG signals.

EMG signal	Ensemble Classifier IDs	Max Kappa Value	First Combiner Majority Voting			Ensemble Classifier IDs	Min Kappa	**Adaptive MV with Average Fixed Rule**			**Adaptive MV with Sugeno Fuzzy Integral**		
			$A_r\%$	$E_r\%$	$CC_r\%$			$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	5 6 7 11 15 16	0.96	97.6	0.7	97.0	4 8 9 10 12 16	-0.18	99.0	1.0	98.0	99.0	1.4	97.6
2	5 6 9 10 11 12	0.98	97.4	0.3	97.1	1 2 5 7 8 9	0.02	99.1	0.6	98.5	98.8	0.6	98.3
3	5 6 7 9 10 11	0.92	96.8	0.8	96.1	1 5 7 9 12 15	-0.02	98.1	0.2	97.8	98.1	0.2	97.8
4	1 2 3 4 13 14	0.92	87.9	1.0	87.0	4 5 7 10 11 14	0.32	93.0	1.0	92.2	93.0	1.7	91.5
5	5 7 9 10 11 12	0.97	99.0	0.8	98.3	1 5 9 10 11 12	-0.08	99.8	0.8	99.0	99.8	0.8	99.0
6	1 2 3 4 15 16	0.94	92.6	0.6	92.1	4 5 7 8 9 11	-0.14	94.9	0.7	94.2	94.4	1.1	93.3
7	5 6 7 8 9 11	0.91	95.3	1.7	93.7	1 4 5 7 11 13	-0.08	97.2	2.3	94.9	96.7	2.9	93.9
8	1 2 3 4 13 14	0.95	92.0	0.3	91.7	1 4 5 7 10 14	-0.11	95.2	1.1	94.1	95.2	1.5	93.8
9	1 5 6 7 9 10	0.90	95.2	0.6	94.6	1 2 5 7 9 10	0.01	97.3	1.2	96.1	97.0	1.8	95.2
10	5 6 7 9 10 11	0.91	94.5	0.9	93.6	5 7 10 11 12 16	0.02	96.5	1.3	95.3	93.9	1.4	92.6
11	4 5 6 7 8 14	0.89	90.6	2.8	88.0	1 4 5 6 7 10	-0.08	95.1	3.3	92.0	94.2	3.7	90.7
12	1 2 3 4 5 6	0.88	89.6	1.3	88.4	1 2 6 7 8 9	-0.13	94.6	1.7	93.0	93.3	2.6	91.0
13	5 7 9 10 11 12	0.84	90.0	4.6	85.8	1 5 7 10 11 12	-0.03	92.4	5.3	87.5	91.5	7.0	85.1
14	5 6 7 8 9 11	0.90	94.4	3.0	91.6	2 5 7 8 9 10	-0.07	96.0	3.5	92.6	95.6	3.8	92.0
15	1 2 3 4 13 14	0.88	80.2	1.4	79.1	1 4 6 9 11 14	-0.11	88.6	2.2	86.6	88.2	2.6	85.9
16	1 2 3 4 5 7	0.85	86.2	0.8	85.5	1 2 5 7 8 9	-0.13	92.6	1.8	90.9	91.5	3.7	88.1
17	1 2 3 4 13 14	0.89	82.5	1.9	80.9	1 2 4 5 6 13	-0.13	88.3	2.3	86.3	88.2	2.8	85.8
18	1 2 3 4 6 16	0.74	69.8	3.9	67.1	2 3 6 11 12 16	-0.08	81.6	6.0	76.7	80.0	6.8	74.5
19	1 2 3 4 15 16	0.80	71.6	1.4	70.6	1 4 7 9 11 16	-0.08	83.1	4.8	79.1	82.2	5.1	78.0
20	1 2 3 4 13 14	0.87	80.4	1.4	79.2	2 3 7 9 11 14	-0.12	87.5	2.2	85.5	87.2	1.8	85.6
		Mean	89.2	1.5	87.9			93.5	2.2	91.5	92.9	2.7	90.5
		MAD	6.6	0.9	6.8			4.1	1.2	4.8	4.1	1.4	5.0

Table 5.23: Performance of 6 out of 16 diversity-based ensemble classifier fusion system using the groups of related simulated EMG signals.

EMG signal	Ensemble Classifier IDs	Max Kappa Value	First Combiner Majority Voting			Ensemble Classifier IDs	Min Kappa	**Adaptive MV with Average Fixed Rule**			**Adaptive MV with Sugeno Fuzzy Integral**		
			$A_r\%$	$E_r\%$	$CC_r\%$			$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1-1	1 2 3 4 5 7	0.90	91.1	1.0	90.2	1 4 7 11 12 16	-0.09	94.5	2.1	92.5	93.0	2.0	91.1
1-2	1 2 3 4 5 7	0.88	90.7	1.3	89.6	1 4 5 7 9 11	-0.10	93.5	2.7	90.9	93.2	3.3	90.1
1-3	1 2 3 4 5 7	0.84	85.6	1.6	84.3	2 3 4 5 7 10	-0.08	91.7	3.4	88.6	91.4	3.9	87.8
1-4	1 2 3 4 5 7	0.89	90.2	1.3	89.1	1 2 4 5 6 7	-0.12	93.4	2.1	91.4	92.8	2.7	90.3
1-5	1 2 3 4 5 7	0.89	88.8	2.6	86.5	1 2 3 5 7 12	-0.10	92.6	3.9	89.0	91.9	4.0	88.2
1-6	2 3 4 6 7 8	0.87	88.6	2.8	86.1	1 2 8 10 11 12	-0.11	95.4	5.3	90.4	95.3	5.7	89.8
1-7	2 4 6 8 10 16	0.76	79.8	2.4	77.9	2 4 6 8 10 16	-0.07	88.4	7.9	81.4	87.9	8.4	80.5
1-8	1 2 4 5 6 8	0.84	86.7	4.2	83.0	4 6 9 11 12 13	-0.07	94.3	7.9	86.8	94.9	7.9	87.5
1-9	1 2 3 4 6 8	0.81	82.8	6.3	77.6	1 4 6 8 9 12	-0.06	91.1	7.9	83.9	92.0	9.2	83.5
2-1	1 3 5 6 7 8	0.84	89.9	2.3	87.9	1 2 5 7 11 12	-0.07	94.0	2.7	91.5	94.0	3.6	90.6
2-2	1 2 3 4 13 14	0.80	76.1	2.7	74.1	1 2 11 12 13 14	-0.09	85.4	4.4	81.7	85.4	5.0	81.2
2-3	1 2 3 4 5 7	0.80	82.1	1.9	80.5	3 4 5 7 9 10	-0.13	90.2	3.9	86.6	88.8	4.2	85.1
2-4	1 2 3 4 5 7	0.82	86.0	2.3	84.0	1 3 5 6 7 12	-0.09	93.9	3.7	90.4	93.9	4.6	89.6
3-1	1 2 3 4 5 7	0.85	86.5	3.6	83.4	2 5 7 10 11 16	-0.11	92.8	5.7	87.5	92.4	6.6	86.2
3-2	1 2 3 4 6 8	0.81	84.1	4.7	80.2	2 3 5 6 7 12	-0.07	94.0	6.8	87.7	93.1	7.2	86.4
3-3	1 2 3 4 5 7	0.76	81.3	5.1	77.2	1 2 5 7 11 14	-0.06	93.3	9.8	84.2	93.2	10.5	83.4
		Mean	85.6	2.9	83.2			92.5	5.0	87.8	92.1	5.6	87.0
		MAD	3.5	1.2	4.0			1.8	2.0	2.8	1.9	2.1	2.8

Table 5.24: Performance of 6 out of 16 diversity-based ensemble classifier fusion system using the real EMG signals.

EMG signal	Ensemble Classifier IDs	Max Kappa Value	First Combiner Majority Voting			Ensemble Classifier IDs	Min Kappa	**Adaptive MV with Average Fixed Rule**			**Adaptive MV with Sugeno Fuzzy Integral**		
			$A_r\%$	$E_r\%$	$CC_r\%$			$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
1	2 4 13 14 15 16	0.88	83.6	1.6	82.3	1 2 4 12 15 16	-0.08	87.9	3.3	85.1	87.3	2.9	84.8
2	2 4 6 8 14 16	0.69	70.8	5.1	67.2	2 6 7 8 9 16	-0.03	90.3	8.4	82.7	89.2	9.6	80.6
3	1 2 3 4 13 14	0.88	79.8	0.5	79.3	1 2 4 7 9 13	-0.12	87.0	1.6	85.7	86.4	2.0	84.6
4	2 4 13 14 15 16	0.91	89.4	0.0	89.4	1 4 9 11 14 16	-0.14	94.3	0.8	93.5	94.3	0.7	93.7
5	2 4 5 7 14 16	0.79	78.4	1.3	77.4	2 4 5 6 7 11	-0.09	89.9	3.3	86.9	90.2	4.3	86.3
		Mean	80.4	1.7	79.1			89.9	3.5	86.8	89.5	3.9	86.0
		MAD	4.9	1.4	5.5			1.9	2.0	2.8	2.2	2.4	3.2

Table 5.25: Summary of experimental results for base classifiers.

Base Classifier	**Independent simulated signals**			**Related simulated signals**			**Real signals**		
	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
e_1	85.2 (5.2)	1.8 (0.9)	83.7 (5.8)	84.0 (3.6)	4.9 (2.2)	79.9 (4.7)	83.9 (4.4)	2.9 (2.7)	81.4 (3.6)
e_2	86.9 (5.2)	1.6 (0.9)	85.5 (5.8)	86.0 (3.7)	5.1 (2.3)	81.6 (4.1)	80.5 (5.8)	4.7 (5.8)	76.7 (6.1)
e_3	86.6 (4.6)	2.5 (1.3)	84.8 (5.4)	86.2 (3.8)	5.7 (2.6)	81.3 (4.2)	82.6 (2.2)	5.9 (5.2)	77.8 (5.8)
e_4	88.6 (4.5)	2.1 (1.3)	86.6 (5.3)	88.3 (3.2)	6.3 (2.7)	82.7 (4.2)	80.6 (5.5)	7.5 (7.1)	74.6 (8.5)
e_5	93.1 (3.0)	4.1 (2.4)	89.3 (4.9)	89.5 (2.5)	8.4 (2.5)	82.1 (3.5)	93.2 (1.6)	6.4 (2.6)	87.3 (3.5)
e_6	96.6 (1.9)	3.2 (1.8)	93.6 (3.5)	91.4 (4.0)	5.6 (1.9)	86.3 (4.2)	91.2 (5.9)	7.1 (5.4)	84.6 (5.7)
e_7	92.0 (3.0)	4.7 (2.4)	87.8 (4.8)	89.8 (2.6)	8.7 (2.3)	82.0 (2.8)	91.4 (1.1)	9.4 (5.3)	82.9 (5.7)
e_8	95.8 (1.9)	3.6 (2.0)	92.5 (3.7)	90.8 (3.3)	6.2 (1.8)	85.2 (3.6)	90.2 (5.1)	8.8 (6.8)	82.2 (9.2)
e_9	93.4 (3.9)	7.4 (3.6)	86.7 (6.8)	88.4 (4.4)	15.5 (6.6)	74.7 (6.2)	90.2 (3.4)	11.1 (5.6)	80.1 (4.3)
e_{10}	94.2 (3.8)	7.1 (3.9)	87.7 (7.0)	86.2 (5.5)	17.1 (7.1)	71.7 (9.8)	88.1 (6.3)	13.4 (5.2)	76.4 (5.9)
e_{11}	92.4 (3.6)	8.6 (4.3)	84.7 (7.0)	88.7 (4.3)	15.5 (5.7)	74.9 (6.7)	92.1 (2.6)	17.4 (6.2)	76.1 (6.8)
e_{12}	93.7 (3.5)	8.4 (4.8)	86.1 (7.5)	86.5 (5.6)	16.4 (7.0)	72.4 (7.8)	90.4 (3.0)	21.5 (10.4)	71.0 (10.1)
e_{13}	86.8 (4.9)	4.8 (2.6)	82.8 (6.0)	85.1 (4.1)	12.2 (4.7)	74.8 (5.4)	86.0 (3.3)	7.9 (6.1)	79.2 (6.5)
e_{14}	88.2 (4.3)	5.6 (3.6)	83.4 (6.4)	83.9 (4.3)	13.3 (5.7)	72.7 (6.6)	83.2 (4.4)	9.5 (8.1)	75.6(10.8)
e_{15}	90.2 (3.8)	6.8 (3.5)	84.3 (6.4)	86.5 (4.0)	12.6 (5.3)	74.8 (5.4)	85.4 (3.6)	9.5 (6.2)	77.5 (7.7)
e_{16}	91.4 (3.2)	7.3 (4.4)	84.9 (6.6)	85.4 (3.2)	14.3 (5.4)	73.2 (6.3)	83.1 (7.1)	15.0 (12.6)	71.6 (16.2)
Mean	91.0	5.0	86.5	87.2	10.5	78.2	87.2	9.9	78.4
MAD	3.0	2.2	2.3	2.0	4.1	4.5	3.8	3.6	3.6

Table 5.26: Summary of experimental results for classifier fusion systems.

Classifier	**Independent simulated signals**			**Related simulated signals**			**Real signals**		
	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$	$A_r\%$	$E_r\%$	$CC_r\%$
Best base	96.6 (1.9)	3.2 (1.8)	93.6 (3.5)	91.4 (4.0)	5.6 (1.9)	86.3 (4.2)	91.2 (5.9)	7.1 (5.4)	84.6 (5.7)
Weakest base	93.7 (3.5)	8.4 (4.8)	86.1 (7.5)	86.5 (5.6)	16.4 (7.0)	72.4 (7.8)	90.4 (3.0)	21.5 (10)	71.0 (10)
Mean base	91.0 (3.0)	5.0 (2.2)	86.5 (2.3)	87.2 (2.0)	10.5 (4.1)	78.2 (4.5)	87.2 (3.8)	9.9 (3.6)	78.4 (3.6)
AMV - SC	90.9 (4.3)	1.5 (1.0)	89.5 (4.9)	85.0 (3.6)	3.0 (1.1)	82.5 (3.9)	82.4 (6.4)	1.9 (1.0)	80.9 (6.8)
AAFR - SC	94.0 (2.7)	3.3 (1.8)	90.9 (4.3)	92.2 (2.0)	5.4 (1.4)	87.2 (2.5)	93.6 (2.1)	4.1 (2.2)	89.8 (3.1)
ASFI - SC	91.7 (3.5)	4.7 (2.1)	87.4 (4.8)	90.6 (2.9)	6.3 (1.1)	84.8 (3.1)	92.1 (2.2)	5.9 (2.4)	86.8 (3.9)
AMVAFR - HC	94.7 (2.7)	2.5 (1.4)	92.4 (3.8)	92.5 (2.0)	4.8 (1.7)	88.1 (2.4)	93.7 (1.7)	3.9 (2.1)	90.1 (2.8)
AMVSFI - HC	94.0 (2.7)	3.3 (1.7)	91.0 (4.1)	91.8 (2.5)	5.5 (1.5)	86.7 (2.7)	92.8 (1.8)	4.8 (2.3)	88.4 (3.3)
AMVAFRD-6/8	94.7 (2.7)	2.2 (1.3)	92.6 (3.5)	92.1 (2.1)	4.6 (1.8)	87.9 (2.4)	92.9 (2.5)	3.8 (2.1)	89.3 (3.0)
AMVSFID-6/8	94.3 (2.8)	2.6 (1.3)	92.0 (3.7)	91.7 (2.3)	4.9 (1.7)	87.2 (2.6)	92.0 (2.5)	4.2 (2.4)	88.1 (2.9)
AMVAFRD-6/16	93.5 (4.1)	2.2 (1.2)	91.5 (4.8)	92.5 (1.8)	5.0 (2.0)	87.8 (2.8)	89.9 (1.9)	3.5 (2.0)	86.8 (2.8)
AMVSFID-6/16	92.9 (4.1)	2.7 (1.4)	90.5 (5.0)	92.1 (1.9)	5.6 (2.1)	87.0 (2.8)	89.5 (2.2)	3.9 (2.4)	86.0 (3.2)

5.4 Discussion of Results and Comparative Study

The performance of the base classifiers used and the performance of the classifier fusion systems can be explained considering the different complexities simulated and real EMG signals. Simulated EMG signals allowed the performance to be evaluated in a controlled way. Signals of known compositions and complexities were presented to each base classifier and an aggregator module combines their output decisions such that the performance in terms of signal characteristics was determined. Real EMG signals that were decomposed manually by a human expert were used as a reference to determine and evaluate performance.

For comparison between classifiers performances, we used the definition of *better performance* stated in Section 5.2.7. The classifier is with better performance, relative to another one, if it is with the highest correct classification rate $CC_r\%$ and lowest error rate $E_r\%$. In situations when there is no clear distinction, we take the difference between the correct classification rate $CC_r\%$ and error rate $E_r\%$ for each classifier and consider the classifier with the higher difference as the one with the better performance.

The used base classifiers passively use firing pattern information to increase or decrease the assignment threshold used to make a MUP assignment. The net affect of these adjustments over the sets of EMG signals studied was an increase in the number of MUP classifications with only a minor increase in the number of incorrect classifications resulting in overall improved and less variable classification rates. The results demonstrate that the adjustments made are appropriate when the IDIs of the MUPTs have small or large variability (CV up to 0.3). However, it is important to realize that these decomposition performance gains are only expected for EMG signals detected during constant or nearly constant force contractions. The increases in the firing pattern consistency statistics used as the number of classification errors increase was studied for MUPTs that had an assumed constant mean IDI and CV. During changing force contractions, the mean IDI and CV of the MUPTs will not be constant. This limitation may restrict the use of the adaptive

base classifiers to research applications. However, for the characterization of the electrophysiological state of a muscle for its clinical assessment the adaptive base classifiers can be very useful.

Considering the fixed classifier ensemble, i.e., the ensemble with same six classifiers $e_1, e_2, e_5, e_6, e_7, e_8$ used across all EMG signals, results across the data sets we see, and as is shown in Table 5.27, that the one level classifier fusion schemes give classification performance better than the average performance of the constituent six classifiers and also better than the performance of the best one except across the independent simulated signals. The hybrid classifier fusion approaches on the other hand give performance that not only exceed the performance of any of the base classifiers forming the ensemble, except across the independent simulated signals, but also reduce the classification errors for all data sets relative to the base classifiers and the one level classifier fusion scheme.

Considering the diversity-based hybrid classifier fusion results, and as shown in Table 5.28, the hybrid classifier fusion approaches provide performance that also exceed the performance of any of the base classifiers forming the ensemble, except across the independent simulated signals, and also reduce the classification errors for all data sets relative to the base classifiers.

Across the set of independent signals of varying complexity and MUP shape variability but with consistent and moderate amounts of IDI variability we see, and as is shown in Tables 5.29 and 5.30, that the one level classifier fusion schemes and the hybrid classifier fusion approaches, both the fixed and diversity-based versions, and in terms of correct classification rate (CC_r) consistently outperformed the base classifiers. The hybrid classifier fusion approach has also reduced the error rate (E_r) relative to the one level classifier fusion schemes except the adaptive majority voting scheme and relative to the base classifiers except the adaptive certainty classifiers that were fed with first-order discrete derivative features. Overall, the hybrid classifier fusion approach also had a lower MAD of the CC_r performance measure than all the base classifiers except the AFNNC base classifiers; and

Table 5.27: Comparison of results of fixed classifier ensemble consisting of six classifiers $e_1, e_2, e_5, e_6, e_7, e_8$.

Classifier	**Independent simulated signals**			**Related simulated signals**			**Real signals**		
	A_r%	E_r%	CC_r%	A_r%	E_r%	CC_r%	A_r%	E_r%	CC_r%
Best base	96.6 (1.9)	3.2 (1.8)	93.6 (3.5)	91.4 (4.0)	5.6 (1.9)	86.3 (4.2)	91.2 (5.9)	7.1 (5.4)	84.6 (5.7)
Weakest base	85.1 (5.2)	1.8 (0.9)	83.7 (5.2)	84.0 (3.6)	4.9 (2.2)	79.9 (4.7)	83.9 (2.8)	2.9 (2.7)	81.4 (3.6)
Mean base	91.6 (3.7)	3.2 (1.0)	88.7 (3.1)	88.6 (2.4)	6.5 (1.4)	82.9 (1.9)	88.4 (4.1)	6.6 (1.9)	82.5 (2.4)
AMV - SC	90.9 (4.3)	1.5 (1.0)	89.5 (4.9)	85.0 (3.6)	3.0 (1.1)	82.5 (3.9)	82.4 (6.4)	1.9 (1.0)	80.9 (6.8)
AAFR - SC	94.0 (2.7)	3.3 (1.8)	90.9 (4.3)	92.2 (2.0)	5.4 (1.4)	87.2 (2.5)	93.6 (2.1)	4.1 (2.2)	89.8 (3.1)
ASFI - SC	91.7 (3.5)	4.7 (2.1)	87.4 (4.8)	90.6 (2.9)	6.3 (1.1)	84.8 (3.1)	92.1 (2.2)	5.9 (2.4)	86.8 (3.9)
AMVAFR - HC	94.7 (2.7)	2.5 (1.4)	92.4 (3.8)	92.5 (2.0)	4.8 (1.7)	88.1 (2.4)	93.7 (1.7)	3.9 (2.1)	90.1 (2.8)
AMVSFI - HC	94.0 (2.7)	3.3 (1.7)	91.0 (4.1)	91.8 (2.5)	5.5 (1.5)	86.7 (2.7)	92.8 (1.8)	4.8 (2.3)	88.4 (3.3)

Table 5.28: Comparison of results of diversity-based classifier ensembles.

Classifier	Independent simulated signals			Related simulated signals			Real signals		
	A_r%	E_r%	CC_r%	A_r%	E_r%	CC_r%	A_r%	E_r%	CC_r%
Best base	96.6 (1.9)	3.2 (1.8)	93.6 (3.5)	91.4 (4.0)	5.6 (1.9)	86.3 (4.2)	91.2 (5.9)	7.1 (5.4)	84.6 (5.7)
Weakest base	93.7 (3.5)	8.4 (4.8)	86.1 (7.5)	86.5 (5.6)	16.4 (7.0)	72.4 (7.8)	90.4 (3.0)	21.5 (10)	71.0 (10)
Mean base	91.0 (3.0)	5.0 (2.2)	86.5 (2.3)	87.2 (2.0)	10.5 (4.1)	78.2 (4.5)	87.2 (3.8)	9.9 (3.6)	78.4 (3.6)
AMVAFRD-6/8	94.7 (2.7)	2.2 (1.3)	92.6 (3.5)	92.1 (2.1)	4.6 (1.8)	87.9 (2.4)	92.9 (2.5)	3.8 (2.1)	89.3 (3.0)
AMVSFID-6/8	94.3 (2.8)	2.6 (1.3)	92.0 (3.7)	91.7 (2.3)	4.9 (1.7)	87.2 (2.6)	92.0 (2.5)	4.2 (2.4)	88.1 (2.9)
AMVAFRD-6/16	93.5 (4.1)	2.2 (1.2)	91.5 (4.8)	92.5 (1.8)	5.0 (2.0)	87.8 (2.8)	89.9 (1.9)	3.5 (2.0)	86.8 (2.8)
AMVSFID-6/16	92.9 (4.1)	2.7 (1.4)	90.5 (5.0)	92.1 (1.9)	5.6 (2.1)	87.0 (2.8)	89.5 (2.2)	3.9 (2.4)	86.0 (3.2)

Table 5.29: Results of fixed ensemble using independent simulated signals.

Classifier	A_r%	E_r%	CC_r%
e_1	85.2 (5.2)	1.8 (0.9)	83.7 (5.8)
e_2	86.9 (5.2)	1.6 (0.9)	85.5 (5.8)
e_5	93.1 (3.0)	4.1 (2.4)	89.3 (4.9)
e_6	96.6 (1.9)	3.2 (1.8)	93.6 (3.5)
e_7	92.0 (3.0)	4.7 (2.4)	87.8 (4.8)
e_8	95.8 (1.9)	3.6 (2.0)	92.5 (3.7)
Mean base	91.6 (3.7)	3.2 (1.0)	88.7 (3.1)
AMV - SC	90.9 (4.3)	1.5 (1.0)	89.5 (4.9)
AAFR - SC	94.0 (2.7)	3.3 (1.8)	90.9 (4.3)
ASFI - SC	91.7 (3.5)	4.7 (2.1)	87.4 (4.8)
AMVAFR - HC	94.7 (2.7)	2.5 (1.4)	92.4 (3.8)
AMVSFI - HC	94.0 (2.7)	3.3 (1.7)	91.0 (4.1)

a lower MAD of E_r except the adaptive majority voting scheme and the adaptive certainty base classifiers. Also on individual signal level, larger performance gains were obtained for signals with greater MUP variability and/or with higher intensity levels (above 80 pps). This suggests that the one level classifier fusion scheme and the hybrid classifier fusion approach has both better and less variable (or more robust) performance.

The groups of related signals allowed performance to be studied as a function of MUP shape and/or IDI variability. Across the three groups of related signals, and as is shown in Tables 5.31 and 5.32, the one level classifier fusion schemes and the hybrid classifier fusion approaches, both the fixed and diversity-based versions, A_r, E_r, and CC_r performances were significantly better than that of base classifiers. This is apparent with either relatively simple signals, such as signal 1-1 which has medium intensity with stable MUP shapes and firing patterns, or for more complex signals, such as signals 1-9 and 3-3 which

Table 5.30: Results of diversity-based ensembles using independent simulated signals.

Classifier	$A_r\%$	$E_r\%$	$CC_r\%$
e_1	85.2 (5.2)	1.8 (0.9)	83.7 (5.8)
e_2	86.9 (5.2)	1.6 (0.9)	85.5 (5.8)
e_3	86.6 (4.6)	2.5 (1.3)	84.8 (5.4)
e_4	88.6 (4.5)	2.1 (1.3)	86.6 (5.3)
e_5	93.1 (3.0)	4.1 (2.4)	89.3 (4.9)
e_6	96.6 (1.9)	3.2 (1.8)	93.6 (3.5)
e_7	92.0 (3.0)	4.7 (2.4)	87.8 (4.8)
e_8	95.8 (1.9)	3.6 (2.0)	92.5 (3.7)
e_9	93.4 (3.9)	7.4 (3.6)	86.7 (6.8)
e_{10}	94.2 (3.8)	7.1 (3.9)	87.7 (7.0)
e_{11}	92.4 (3.6)	8.6 (4.3)	84.7 (7.0)
e_{12}	93.7 (3.5)	8.4 (4.8)	86.1 (7.5)
e_{13}	86.8 (4.9)	4.8 (2.6)	82.8 (6.0)
e_{14}	88.2 (4.3)	5.6 (3.6)	83.4 (6.4)
e_{15}	90.2 (3.8)	6.8 (3.5)	84.3 (6.4)
e_{16}	91.4 (3.2)	7.3 (4.4)	84.9 (6.6)
Mean base	91.0 (3.0)	5.0 (2.2)	86.5 (2.3)
AMVAFRD-6/8	94.7 (2.7)	2.2 (1.3)	92.6 (3.5)
AMVSFID-6/8	94.3 (2.8)	2.6 (1.3)	92.0 (3.7)
AMVAFRD-6/16	93.5 (4.1)	2.2 (1.2)	91.5 (4.8)
AMVSFID-6/16	92.9 (4.1)	2.7 (1.4)	90.5 (5.0)

Table 5.31: Results of fixed ensemble using related simulated signals.

Classifier	A_r%	E_r%	CC_r%
e_1	84.0 (3.6)	4.9 (2.2)	79.9 (4.7)
e_2	86.0 (3.7)	5.1 (2.3)	81.6 (4.1)
e_5	89.5 (2.5)	8.4 (2.5)	82.1 (3.5)
e_6	91.4 (4.0)	5.6 (1.9)	86.3 (4.2)
e_7	89.8 (2.6)	8.7 (2.3)	82.0 (2.8)
e_8	90.8 (3.3)	6.2 (1.8)	85.2 (3.6)
Mean base	88.6 (2.4)	6.5 (1.4)	82.9 (1.9)
AMV - SC	85.0 (3.6)	3.0 (1.1)	82.5 (3.9)
AAFR - SC	92.2 (2.0)	5.4 (1.4)	87.2 (2.5)
ASFI - SC	90.6 (2.9)	6.3 (1.1)	84.8 (3.1)
AMVAFR - HC	92.5 (2.0)	4.8 (1.7)	88.1 (2.4)
AMVSFI - HC	91.8 (2.5)	5.5 (1.5)	86.7 (2.7)

has high intensity with unstable MUP shapes and firing patterns, where the classifier fusion approaches provided a significant improvement in CC_r while reducing E_r except the adaptive majority voting scheme which is not a suitable candidate for comparison due to its lower assignment rate A_r. In addition, the variability of the performance of the classifier fusion approaches were less than all the base classifiers except the adaptive majority voting which exhibits a higher MAD in A_r, and CC_r and lower MAD in E_r.

For the real EMG signals studied, the one level classifier fusion scheme and the hybrid classifier fusion approach, both the fixed and diversity-based versions whose results were shown in Tables 5.33 and 5.34, respectively, had improved classification performance in terms of CC_r relative to all base classifiers and also reduced the error rate except for the adaptive certainty base classifier e_1. The variability of the performance of the hybrid classifier fusion approaches was less than all the base classifiers except classifier e_1.

Table 5.32: Results of diversity-based ensembles using related simulated signals.

Classifier	$A_r\%$	$E_r\%$	$CC_r\%$
e_1	84.0 (3.6)	4.9 (2.2)	79.9 (4.7)
e_2	86.0 (3.7)	5.1 (2.3)	81.6 (4.1)
e_3	86.2 (3.8)	5.7 (2.6)	81.3 (4.2)
e_4	88.3 (3.2)	6.3 (2.7)	82.7 (4.2)
e_5	89.5 (2.5)	8.4 (2.5)	82.1 (3.5)
e_6	91.4 (4.0)	5.6 (1.9)	86.3 (4.2)
e_7	89.8 (2.6)	8.7 (2.3)	82.0 (2.8)
e_8	90.8 (3.3)	6.2 (1.8)	85.2 (3.6)
e_9	88.4 (4.4)	15.5 (6.6)	74.7 (6.2)
e_{10}	86.2 (5.5)	17.1 (7.1)	71.7 (9.8)
e_{11}	88.7 (4.3)	15.5 (5.7)	74.9 (6.7)
e_{12}	86.5 (5.6)	16.4 (7.0)	72.4 (7.8)
e_{13}	85.1 (4.1)	12.2 (4.7)	74.8 (5.4)
e_{14}	83.9 (4.3)	13.3 (5.7)	72.7 (6.6)
e_{15}	86.5 (4.0)	12.6 (5.3)	74.8 (5.4)
e_{16}	85.4 (3.2)	14.3 (5.4)	73.2 (6.3)
Mean base	87.2 (2.0)	10.5 (4.1)	78.2 (4.5)
AMVAFRD-6/8	92.1 (2.1)	4.6 (1.8)	87.9 (2.4)
AMVSFID-6/8	91.7 (2.3)	4.9 (1.7)	87.2 (2.6)
AMVAFRD-6/16	92.5 (1.8)	5.0 (2.0)	87.8 (2.8)
AMVSFID-6/16	92.1 (1.9)	5.6 (2.1)	87.0 (2.8)

Table 5.33: Results of fixed ensemble using real signals

Classifier	$A_r\%$	$E_r\%$	$CC_r\%$
e_1	83.9 (4.4)	2.9 (2.7)	81.4 (3.6)
e_2	80.5 (5.8)	4.7 (5.8)	76.7 (6.1)
e_5	93.2 (1.6)	6.4 (2.6)	87.3 (3.5)
e_6	91.2 (5.9)	7.1 (5.4)	84.6 (5.7)
e_7	91.4 (1.1)	9.4 (5.3)	82.9 (5.7)
e_8	90.2 (5.1)	8.8 (6.8)	82.2 (9.2)
Mean base	88.4 (4.1)	6.6 (1.9)	82.5 (2.4)
AMV - SC	82.4 (6.4)	1.9 (1.0)	80.9 (6.8)
AAFR - SC	93.6 (2.1)	4.1 (2.2)	89.8 (3.1)
ASFI - SC	92.1 (2.2)	5.9 (2.4)	86.8 (3.9)
AMVAFR - HC	93.7 (1.7)	3.9 (2.1)	90.1 (2.8)
AMVSFI - HC	92.8 (1.8)	4.8 (2.3)	88.4 (3.3)

Table 5.34: Results of diversity-based ensembles using real signals.

Classifier	$A_r\%$	$E_r\%$	$CC_r\%$
e_1	83.9 (4.4)	2.9 (2.7)	81.4 (3.6)
e_2	80.5 (5.8)	4.7 (5.8)	76.7 (6.1)
e_3	82.6 (2.2)	5.9 (5.2)	77.8 (5.8)
e_4	80.6 (5.5)	7.5 (7.1)	74.6 (8.5)
e_5	93.2 (1.6)	6.4 (2.6)	87.3 (3.5)
e_6	91.2 (5.9)	7.1 (5.4)	84.6 (5.7)
e_7	91.4 (1.1)	9.4 (5.3)	82.9 (5.7)
e_8	90.2 (5.1)	8.8 (6.8)	82.2 (9.2)
e_9	90.2 (3.4)	11.1 (5.6)	80.1 (4.3)
e_{10}	88.1 (6.3)	13.4 (5.2)	76.4 (5.9)
e_{11}	92.1 (2.6)	17.4 (6.2)	76.1 (6.8)
e_{12}	90.4 (3.0)	21.5 (10.4)	71.0 (10.1)
e_{13}	86.0 (3.3)	7.9 (6.1)	79.2 (6.5)
e_{14}	83.2 (4.4)	9.5 (8.1)	75.6(10.8)
e_{15}	85.4 (3.6)	9.5 (6.2)	77.5 (7.7)
e_{16}	83.1 (7.1)	15.0 (12.6)	71.6 (16.2)
Mean base	87.2 (3.8)	9.9 (3.6)	78.4 (3.6)
AMVAFRD-6/8	92.9 (2.5)	3.8 (2.1)	89.3 (3.0)
AMVSFID-6/8	92.0 (2.5)	4.2 (2.4)	88.1 (2.9)
AMVAFRD-6/16	89.9 (1.9)	3.5 (2.0)	86.8 (2.8)
AMVSFID-6/16	89.5 (2.2)	3.9 (2.4)	86.0 (3.2)

Figures 5.2, 5.3, 5.4, and 5.5 show the decomposition results of base classifiers ACC, AFNNC, ANCCC, and ApCC for related EMG signal 2-3, respectively. Figure 5.6 presents decomposition results for related EMG signal 2-3 using the fixed ensemble hybrid classifier fusion approach consisting of majority voting and average rule aggregator. Signal 2-3 has 6 MUPTs and was simulated to have a jitter value of 150 μs and an IDI CV of 0.15. The effect of the MUP shape and motor unit firing pattern variability can be seen in the shimmer, and IDI histogram and MUP firing time plots, respectively. The hybrid classifier fusion system was able to sufficiently track each MUPT such that accurate estimates of the mean IDI and the standard deviation of the IDIs for each train were obtained by the EFE algorithm. When comparing with decomposition results of the base classifiers we see that the fixed ensemble hybrid classifier fusion system decomposition is better than that for the ACC in terms of the identification rate for each MUPT and close to the AFNNC, which is the best base classifier in the fixed ensemble. Specifically, if we look at the firing pattern of MUPT #4, both base classifiers, i.e., the ACC and AFNNC, recognize a somehow sparse train due to the similarity with MUPT #2 and MUPT #1 which is apparent in the MUP template plot, where the identification rate for the ACC = 60% and for the AFNNC = 61% but for the fixed ensemble adaptive hybrid classifier fusion system it is 76%. This effect demonstrates the complementary act of the base classifiers in the ensemble to correct the errors. Now when comparing with the firing pattern of MUPT #4 generated by the two matched template filter classifiers, i.e., the ANCCC and the ApCC, we see the identification rate for ANCCC = 89% and for ApCC = 85% but this is not reliable as both shows for the signal 2-3 a high error rate, which is 15.8% for ANCCC and 9.0% for ApCC.

Figure 5.7 presents decomposition results for related EMG signal 2-3 using the 6 out of 8 diversity-based ensembles adaptive hybrid classifier fusion approach consisting of majority voting and average fixed rule aggregator. From Figure 5.7, we notice that the decomposition results is close to that for the fixed fixed ensemble adaptive hybrid classifier fusion approach shown in Figure 5.6.

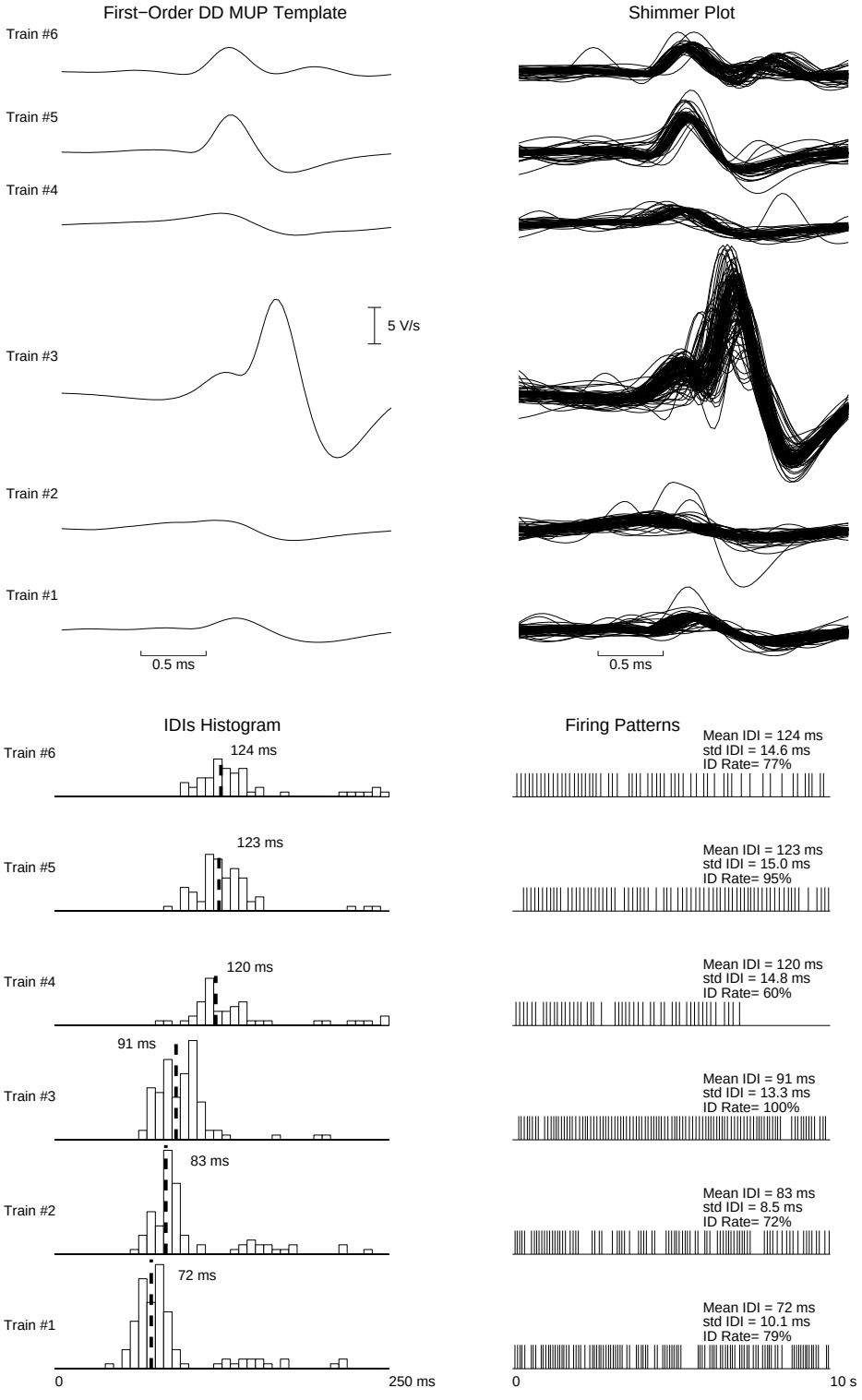

Figure 5.2: Signal 2-3 decomposition summary using base classifier ACC with sequential MUPs seeding.

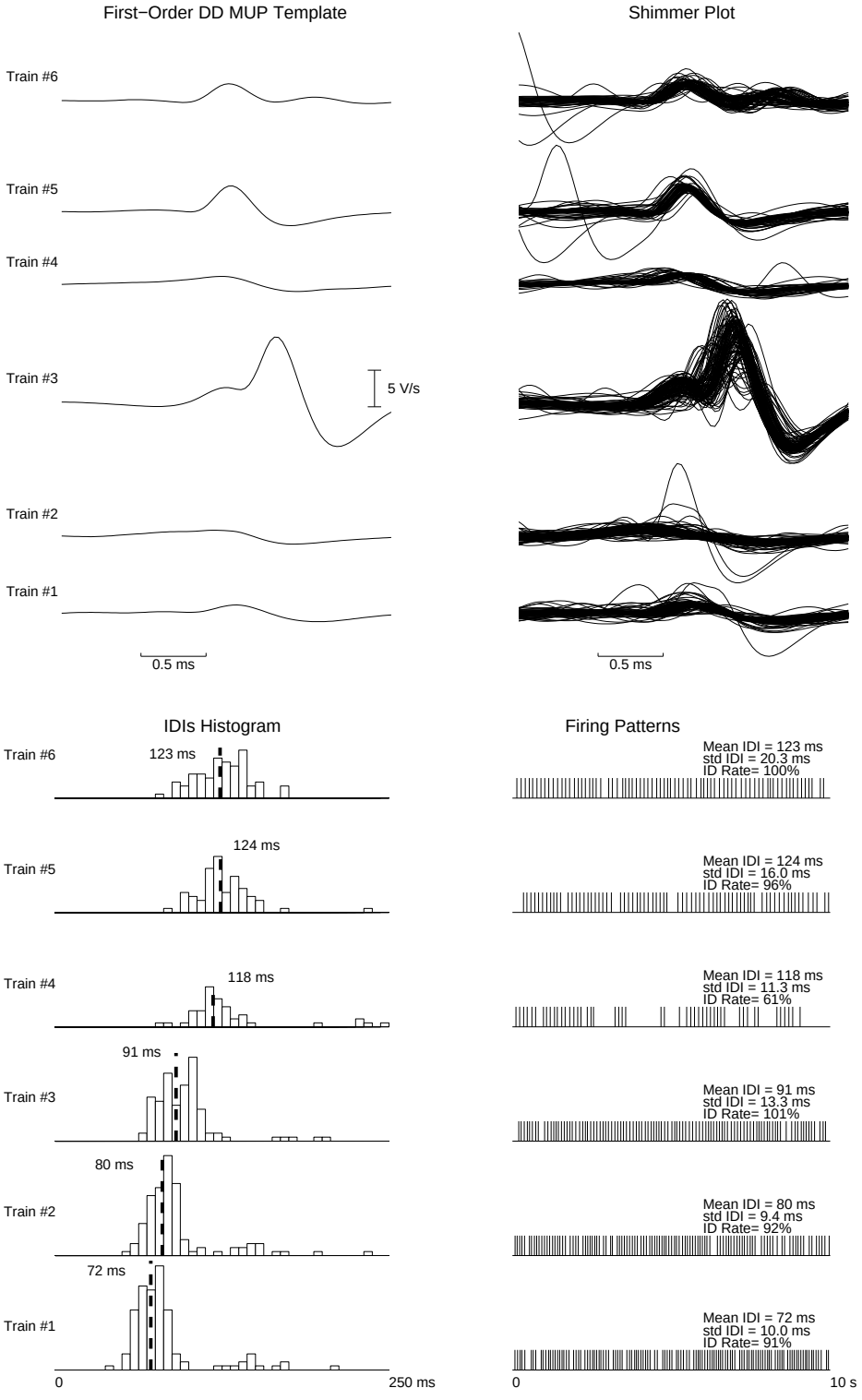

Figure 5.3: Signal 2-3 decomposition summary using base classifier AFNNC with sequential MUPs seeding.

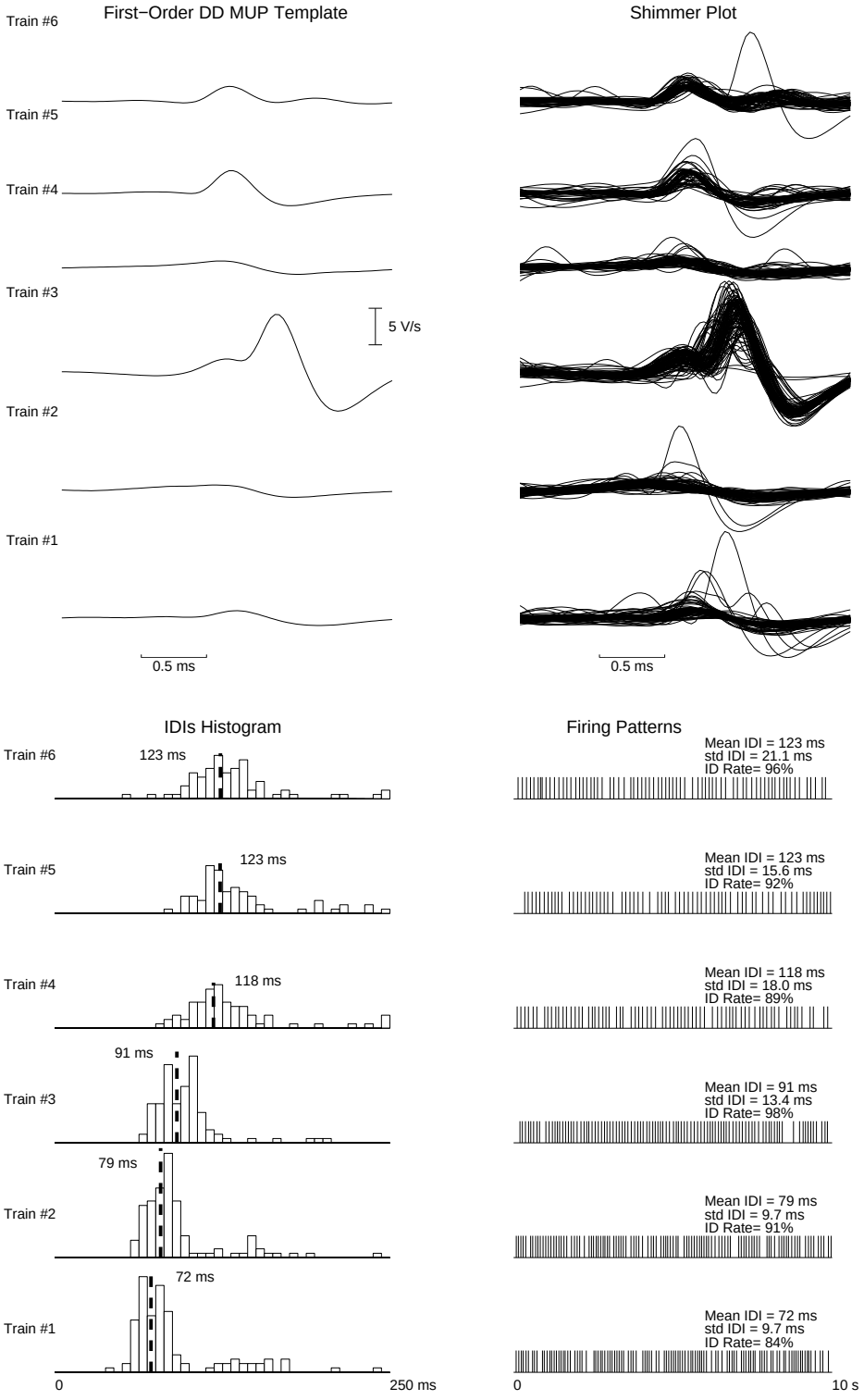

Figure 5.4: Signal 2-3 decomposition summary using base classifier ANCCC with sequential MUPs seeding.

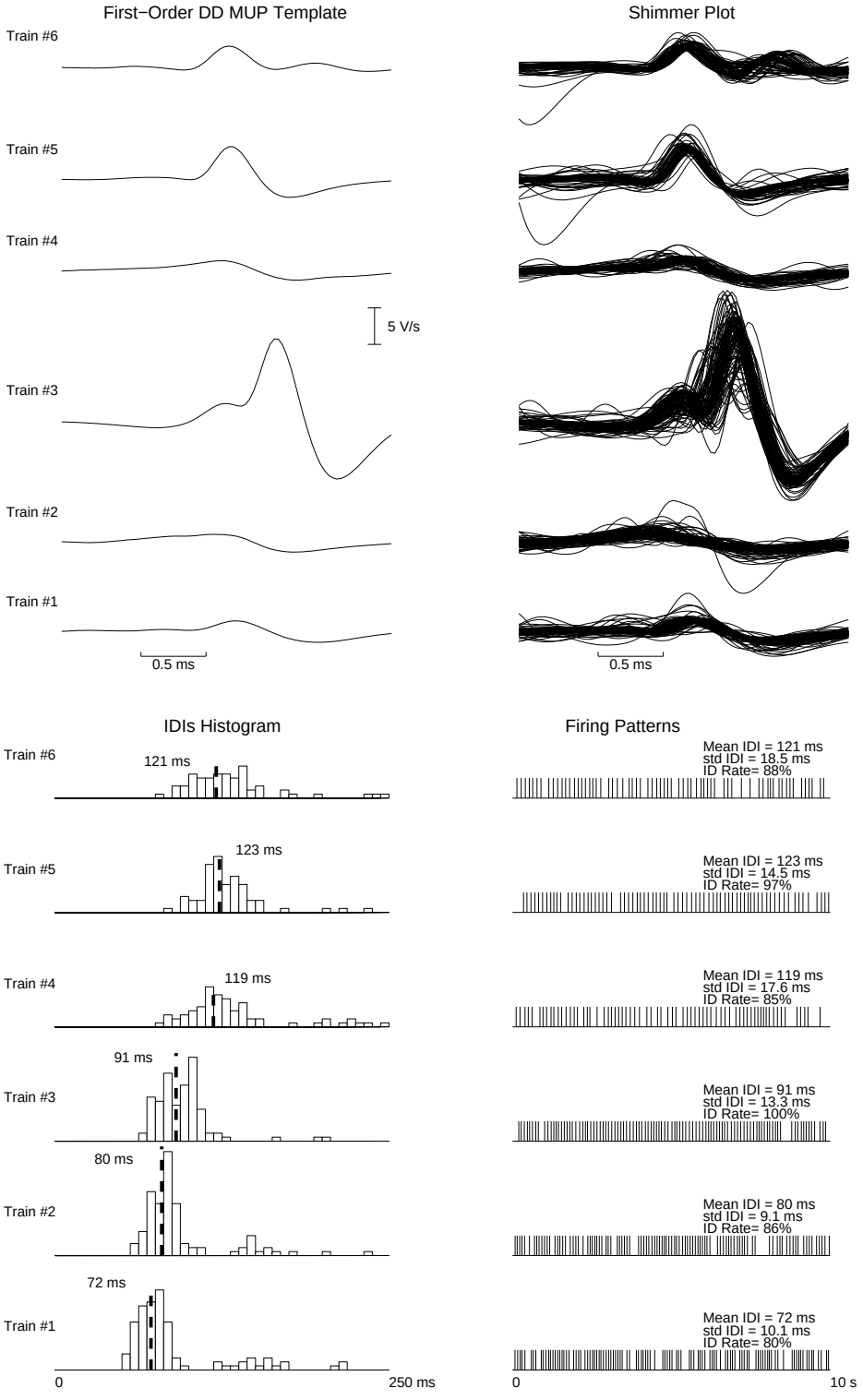

Figure 5.5: Signal 2-3 decomposition summary using base classifier ApCC with sequential MUPs seeding.

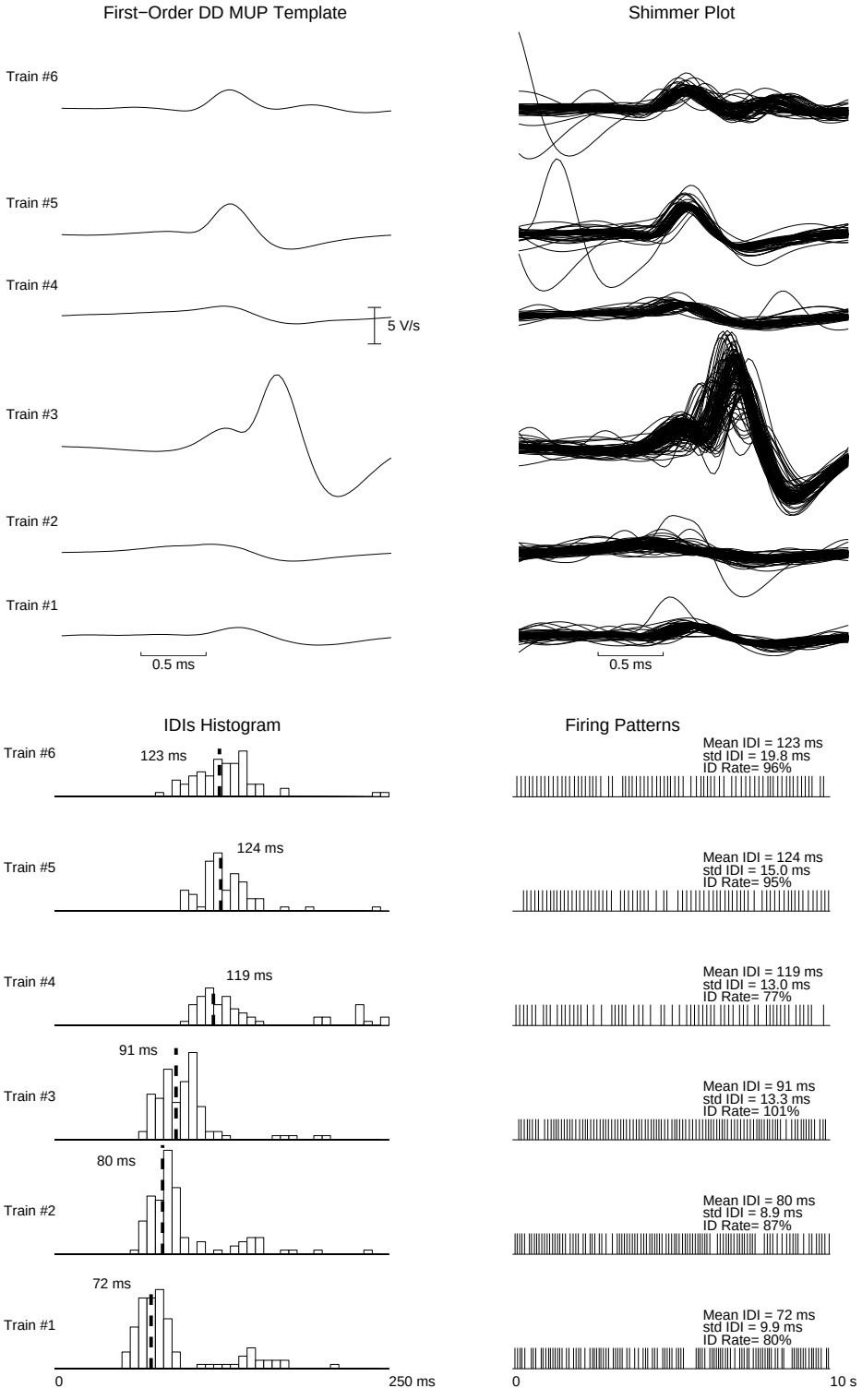

Figure 5.6: Signal 2-3 decomposition summary using the fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting and average fixed rule aggregator.

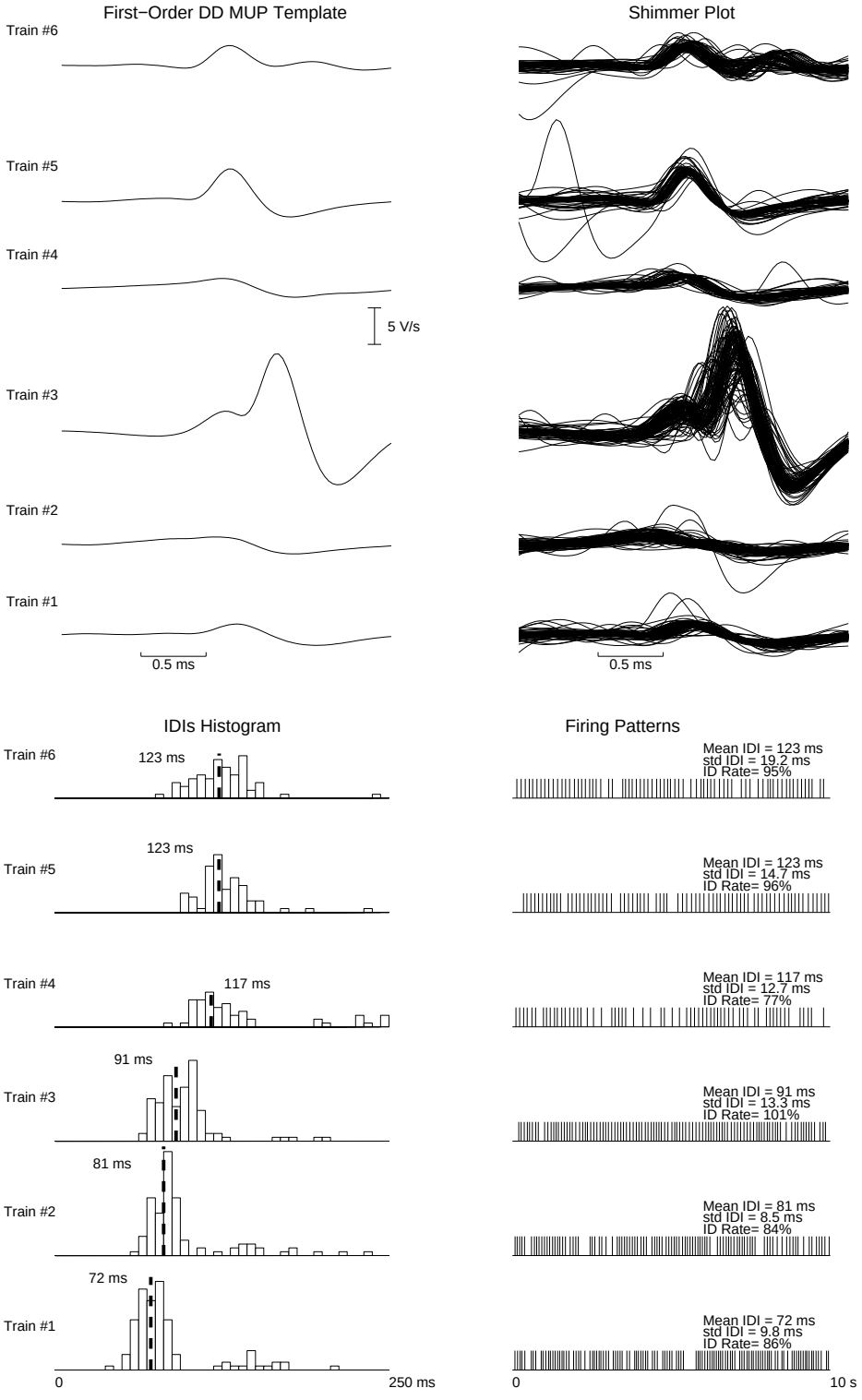

Figure 5.7: Signal 2-3 decomposition summary using the 6 out of 8 diversity-based ensembles hybrid classifier fusion consisting of majority voting and average rule aggregator.

The complementary act of the base classifiers in the ensemble is also demonstrated in Figure 5.12, which displays a 1 s interval of the decomposition results for the 6 MUPTs of signal 2-3 and the unassigned MUPs as decomposed by the fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting and average fixed rule aggregator. Portions of MUPTs are displayed with the time scale used for displaying MUPs expanded by a factor of 10 relative to the time scale used to depict the firing times. This allows the actual shape of each MUP to be better visualized. An erroneous MUP classification is indicated by displaying the number of the correct train next to the MUP. Most of the errors made are related to the shape variability of MUPs occurring at expected firing times for other trains. In these cases, the information provided by the MUP shape and the firing pattern information is not sufficient to make a correct decision. When comparing with MUP traces of the base classifiers ACC, AFNNC, ANCCC, and ApCC shown in Figures 5.8, 5.9, 5.10, and 5.11, respectively, we see that the fixed ensemble adaptive hybrid classifier fusion system shows a MUP traces with no error and only five MUPs unassigned. The complementary effect is evident when the fixed ensemble adaptive hybrid classifier fusion system detected and corrected the error shown in MUPT #4 and clearly apparent in Figure 5.9, which is misclassified by AFNNC, ApCC and correctly classified by ACC, ANCCC, and switched the erroneous MUP belonging to MUPT #2 with the correct MUP belonging to MUPT #4.

Figure 5.13 shows another improvement in performance using the 6 out of 8 diversity-based adaptive hybrid classifier fusion approach. The two ensembles in the system works in a complementary manner such that they reduces the number of unassigned MUPs to one only and also with no erroneous assignment MUPs.

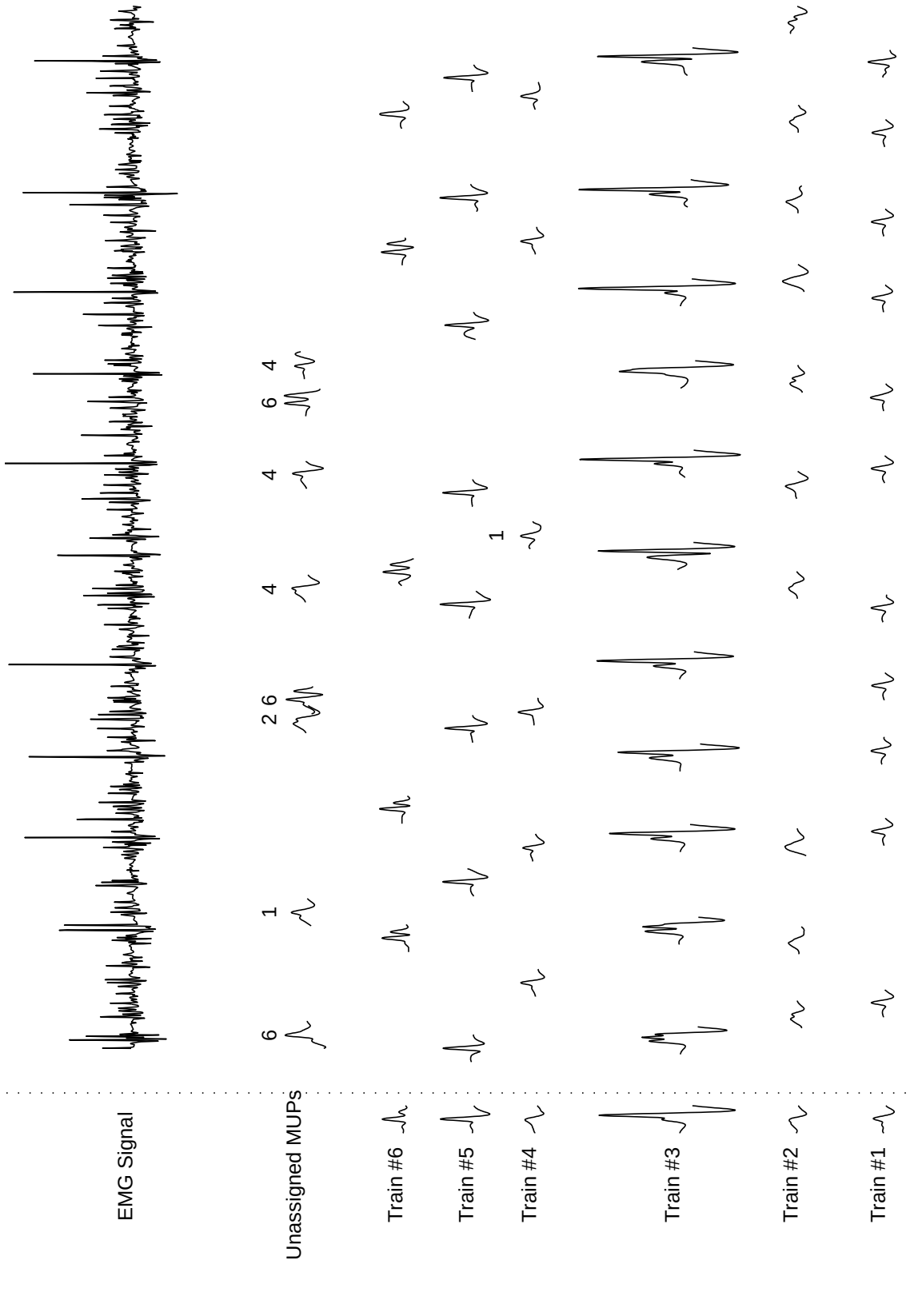

Figure 5.8: MUP traces for a 1 s from 6 s to 7 s interval decomposition results of signal 2-3 using base classifier ACC.

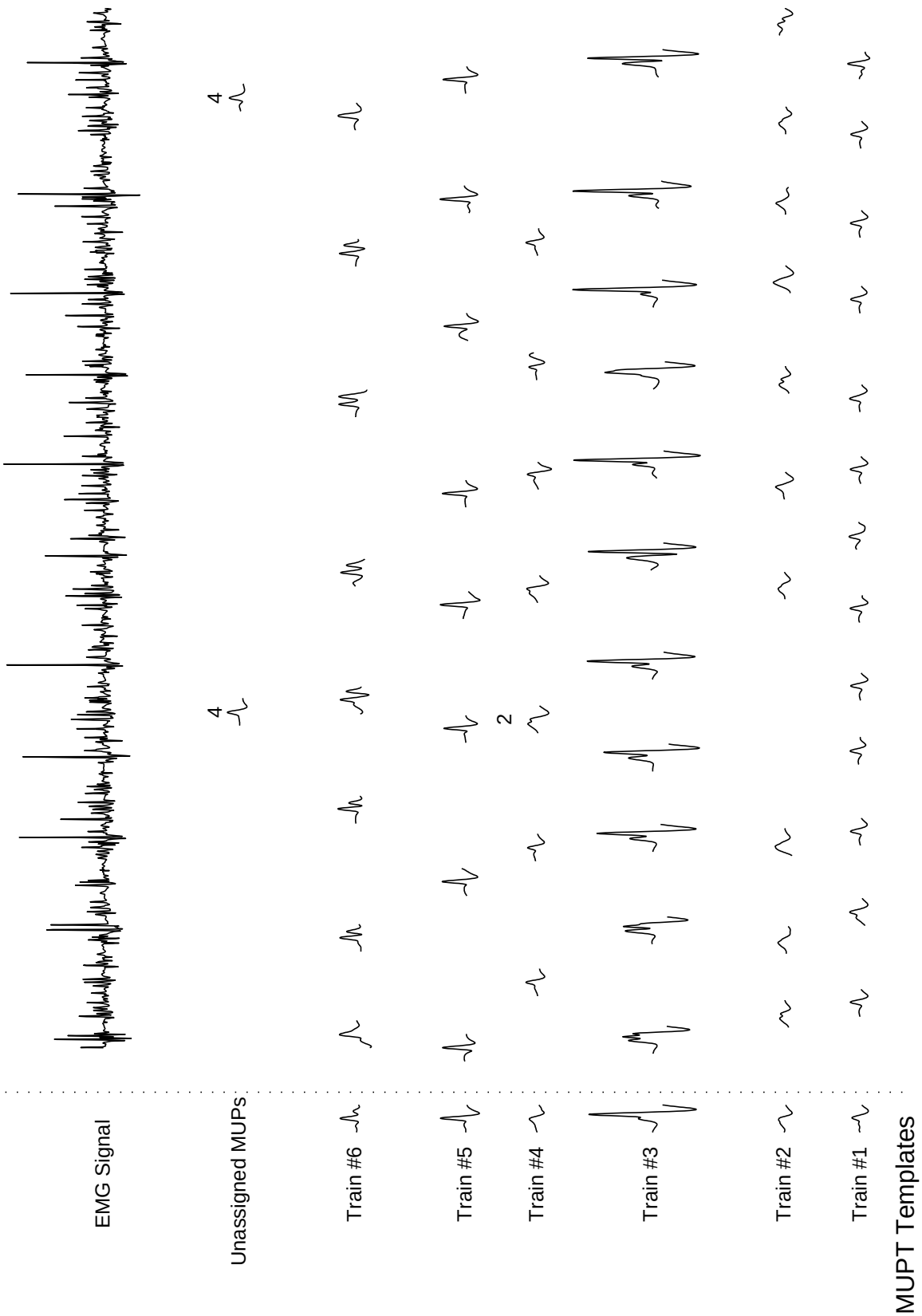

Figure 5.9: MUP traces for a 1 s from 6 s to 7 s interval decomposition results of signal 2-3 using base classifier AFNNC.

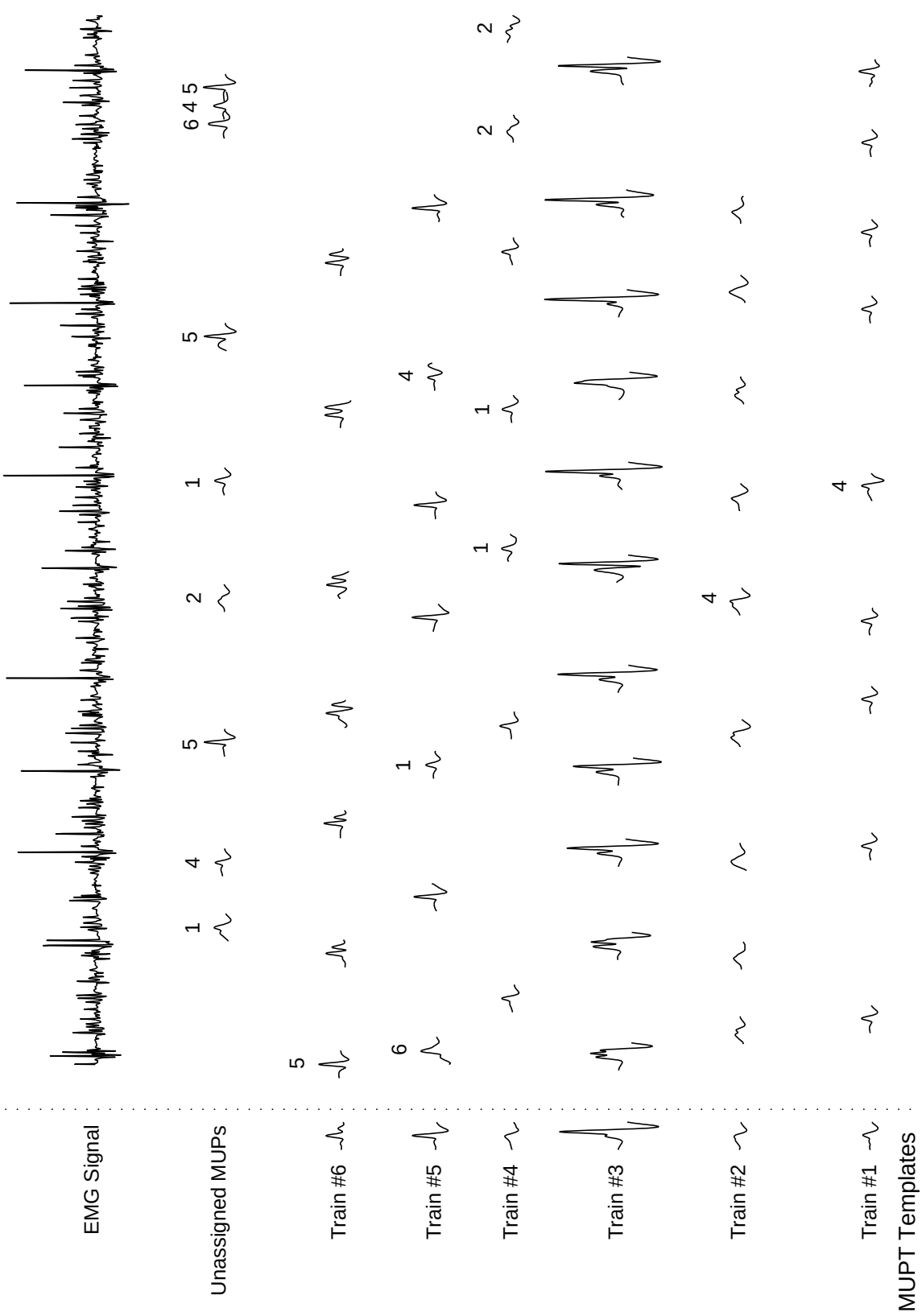

Figure 5.10: MUP traces for a 1 s from 6 s to 7 s interval decomposition results of signal 2-3 using base classifier ANCCC.

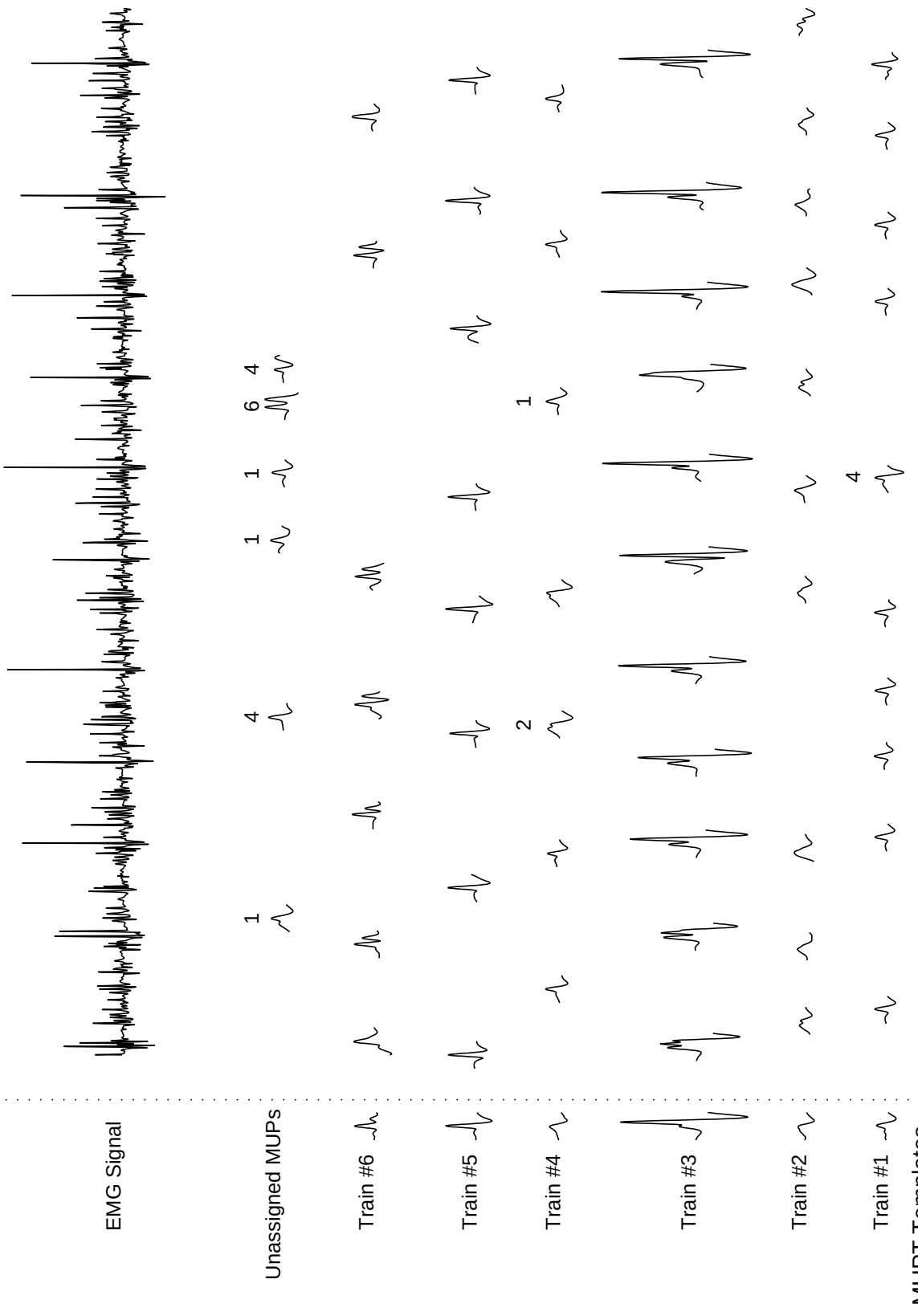

Figure 5.11: MUP traces for a 1 s from 6 s to 7 s interval decomposition results of signal 2-3 using base classifier ApCC.

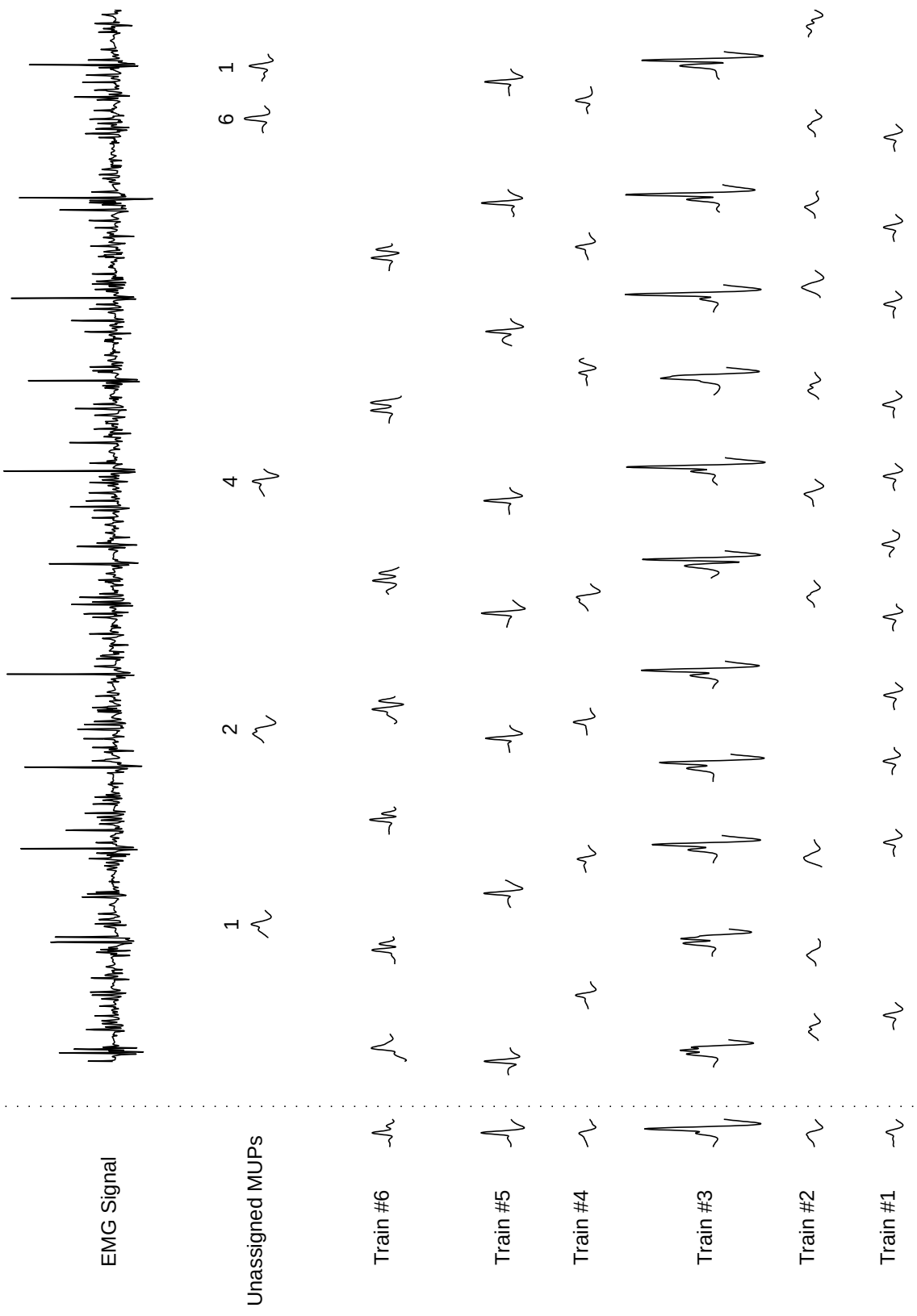

Figure 5.12: MUP traces for a 1 s from 6 s to 7 s interval decomposition results of signal 2-3 using the fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting and average fixed rule aggregator.

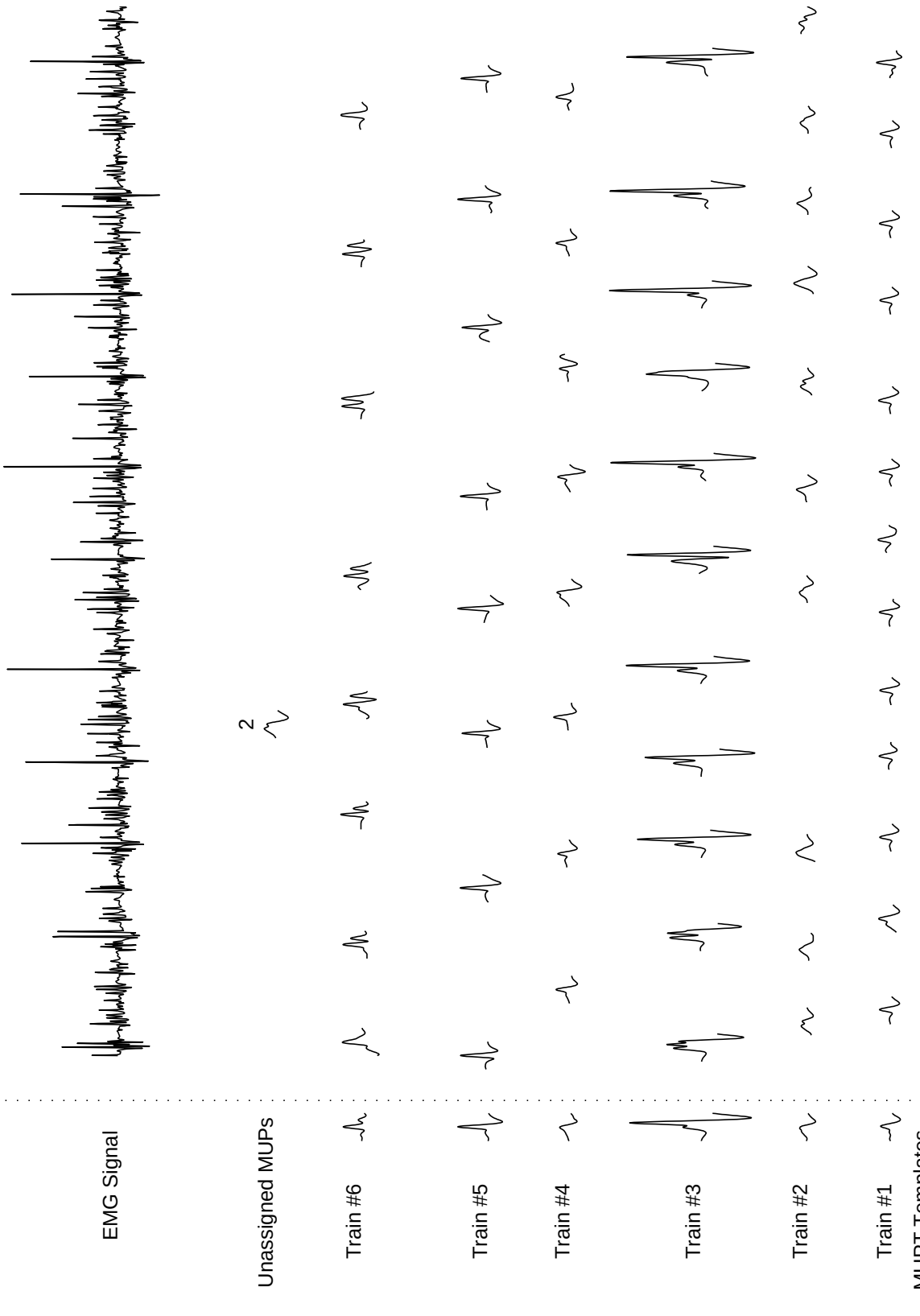

Figure 5.13: MUP traces for a 1 s from 6 s to 7 s interval decomposition results of signal 2-3 using the 6 out of 8 diversity-based adaptive hybrid classifier fusion approach consisting of majority voting and average fixed rule aggregator.

On the other hand, Figure 5.15 presents decomposition results for the related EMG signal 2-4 using the fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting and Sugeno fuzzy integral aggregator. Signal 2-4 has 6 MUPTs and was simulated to have a jitter value of 150 μs and an IDI CV of 0.3. The fixed ensemble adaptive hybrid classifier fusion decomposition of signal 2-4 is more complete (CC_r = 89.4%) and has fewer errors (E_r = 5.6%) compared to the best base classifier (CC_r = 86.4%, E_r = 3.9%). When comparing with the decomposition result of base classifier ACC shown in Figure 5.14, we see that the fixed ensemble adaptive hybrid classifier fusion system decomposition is better than that for the ACC in terms of the identification rate for each MUPT. Specifically, if we look at the firing pattern of MUPT #4, the ACC results are somewhat sparse with an identification rate of 58% but for the fixed ensemble adaptive hybrid classifier fusion system the identification rate is 92%. Similarly, the increase in the identification rate also occurred for MUPTs #1, #2, #5, and #6. Again, this effect is a result of the base classifiers in the ensemble acting as a team to correct the errors.

The team work of the base classifiers is also shown in Figure 5.17, which displays a 1 s interval of the decomposition results for the 6 MUPTs of signal 2-4 and the unassigned MUPs as decomposed by the fixed ensemble adaptive hybrid classifier fusion approach. When compared with classified MUPs of the ACC base classifier shown in 5.16, we see that the fixed ensemble adaptive hybrid classifier fusion system makes only six erroneous assignments and shows only two MUPs unassigned for this epoch of time compared to four erroneous assignments and nine MUPs left unassigned by the ACC. The complementary effect of the team members is again evident when it is noticed that the fixed ensemble adaptive hybrid classifier fusion system detected and corrected some errors, but when it tried to reassign three of the MUPs it erroneously assigned them.

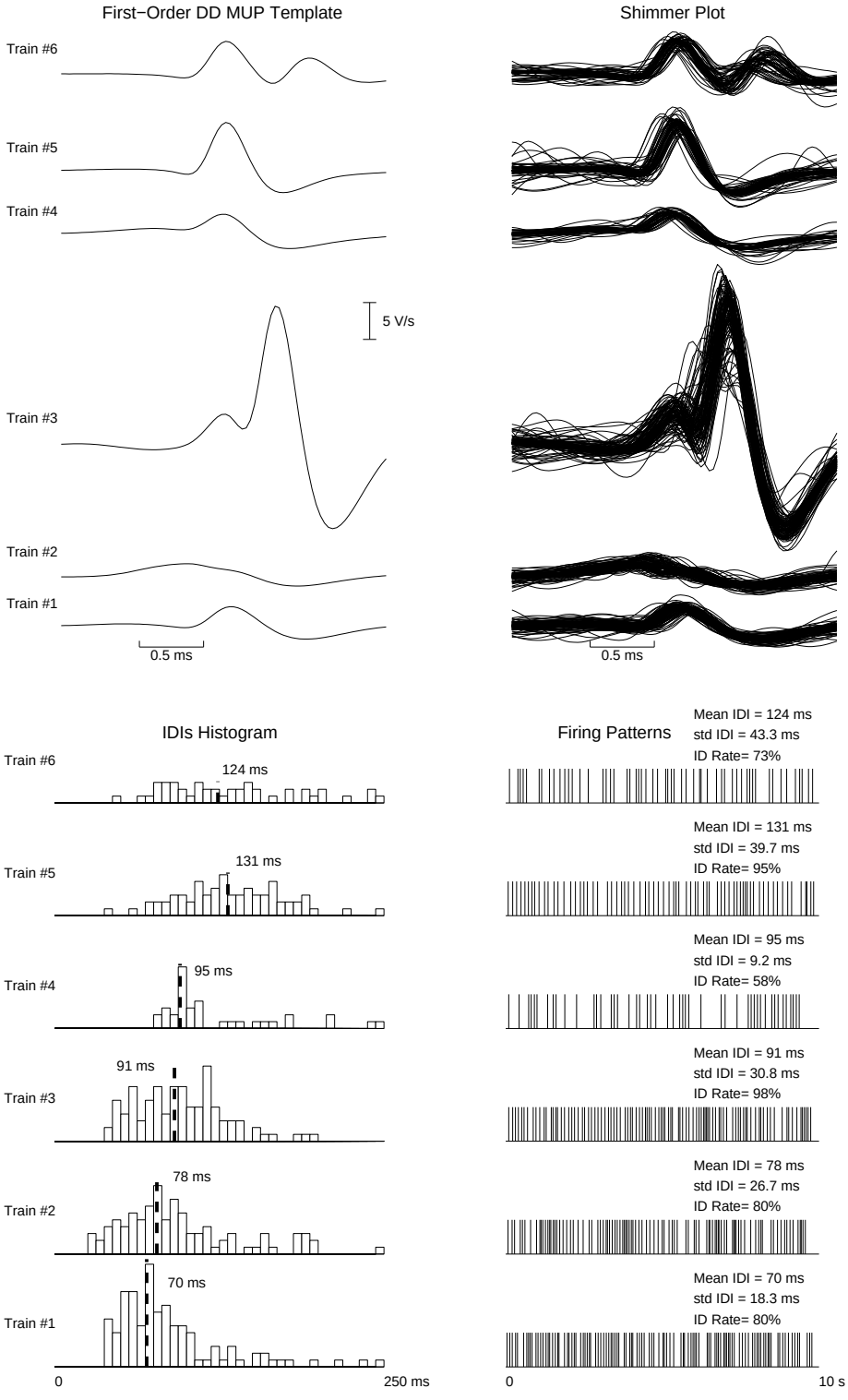

Figure 5.14: Signal 2-4 decomposition summary using base classifier ACC with high certainty MUPs seeding.

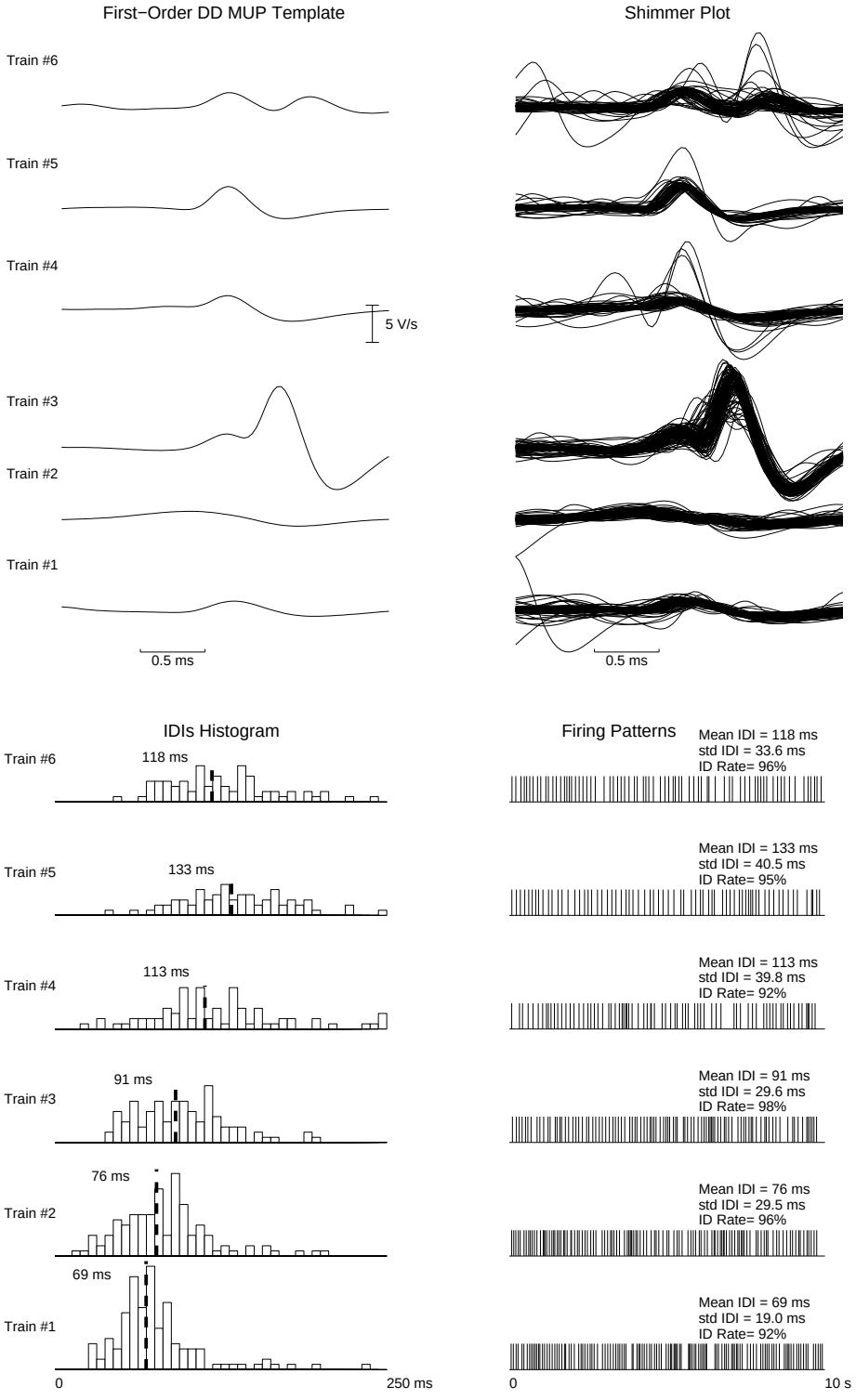

Figure 5.15: Signal 2-4 decomposition summary using fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting and Sugeno fuzzy integral aggregator.

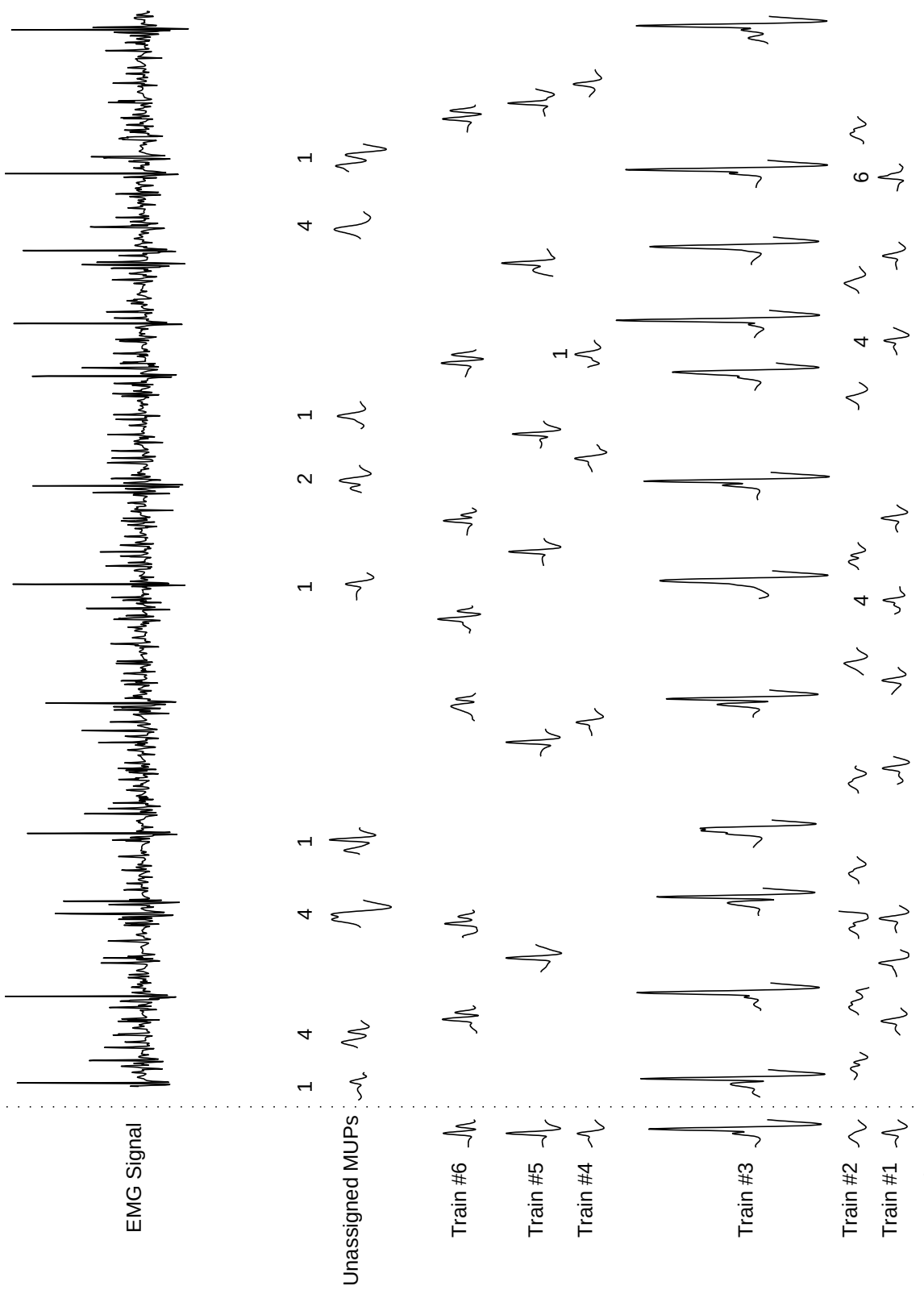

Figure 5.16: MUP traces for a 1 s from 4 s to 5 s interval decomposition results of signal 2-4 using base classifier ACC.

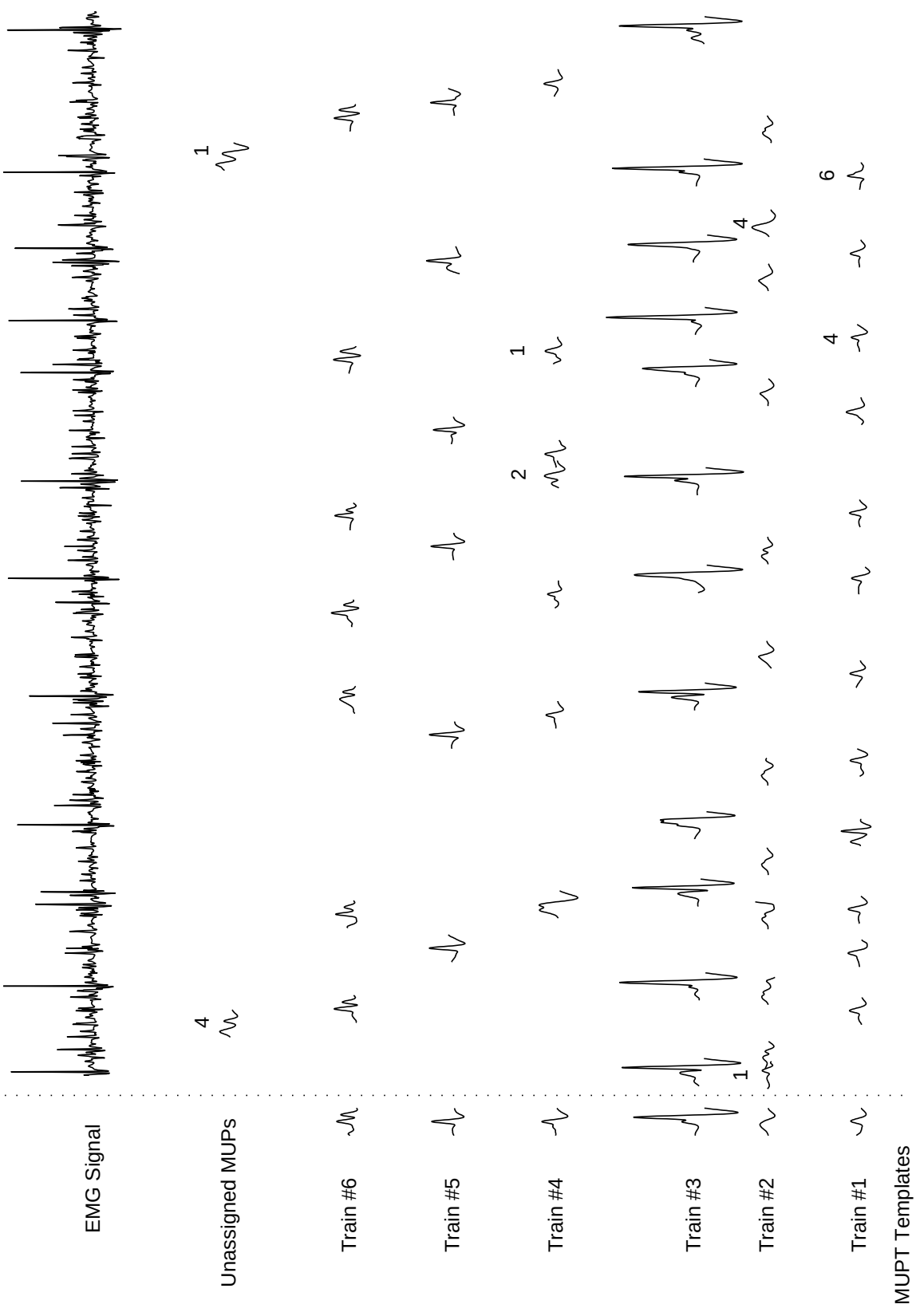

Figure 5.17: MUP traces for a 1 s from 4 s to 5 s interval decomposition results of signal 2-4 using the fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting and Sugeno fuzzy integral aggregator.

Figures 5.18 presents decomposition results for the related EMG signal 2-4 using the 6 out of 16 diversity-based ensembles adaptive hybrid classifier fusion approach consisting of majority voting and Sugeno fuzzy integral aggregator. From Figure 5.18, we notice that the decomposition results is close to that for the fixed ensemble adaptive hybrid classifier fusion approach shown in Figure 5.15. The team work of the base classifiers is also shown in Figures 5.19 and 5.20, which displays a 1 s interval of the decomposition results for the 6 MUPTs of signal 2-4 and the unassigned MUPs as decomposed by the 6 out of 8 and 6 out of 16 diversity-based ensembles, respectively. When compared with classified MUPs of the ACC base classifier shown in 5.16 and the fixed ensemble hybrid fusion system shown in 5.17, we see that the 6 out of 8 diversity-based ensembles makes only three erroneous assignments and shows only five MUPs unassigned for this epoch of time, but the 6 out of 16 diversity-based ensembles makes six erroneous assignments and shows five MUPs unassigned.

The weaker performance of the 6 out of 16 diversity-based ensembles relative to the 6 out of 8 diversity-based ensembles, when decomposing the related EMG signal 2-4, is mainly due to the fact that the ensemble at the second stage of the diversity-based hybrid fusion system had joined in its team classifier e_{12}, refer to Table 5.23, which has poor performance ($A_r = 77.7\%$, $E_r = 10.0\%$, and $CC_r = 69.9\%$).

The fuzzy densities $g^{i/k}$ play an essential role through the determination of the fuzzy measures used in the fusion process of the fuzzy integral method. In reality when dealing with real EMG signals, there is no a priori information related to the signal acquired. This makes the estimation of fuzzy densities $g^{i/k}$ in terms of the base classifier recognition accuracies impossible. Therefore, there is need for employing other approaches, the densities $g^{i/k}$ values may be estimated either using an adaptive fuzzy integral [10], [79] or using genetic algorithms to search for an optimal set of $g^{i/k}$ values [80], [112].

Based on the summary of experimental results shown in Table 5.26, we notice that the fixed ensemble adaptive hybrid classifier fusion approach consisting of majority voting

and average fixed rule aggregator has the best performance across the used data sets. This result is apparent due to the manual selection of ensemble classifiers done in the experiment, where we chose the best six from the pool of classifiers. We get somehow close results to the best by using the automatic selection based on the diversity measure. Table 5.26 shows that the 6 out of 8 diversity-based ensembles adaptive hybrid classifier fusion approach consisting of majority voting and average fixed rule aggregator has performance close to the best. This performance shows the potential of the automatic selection of ensemble classifiers based on assessing the level of agreement between the classifiers in the ensemble.

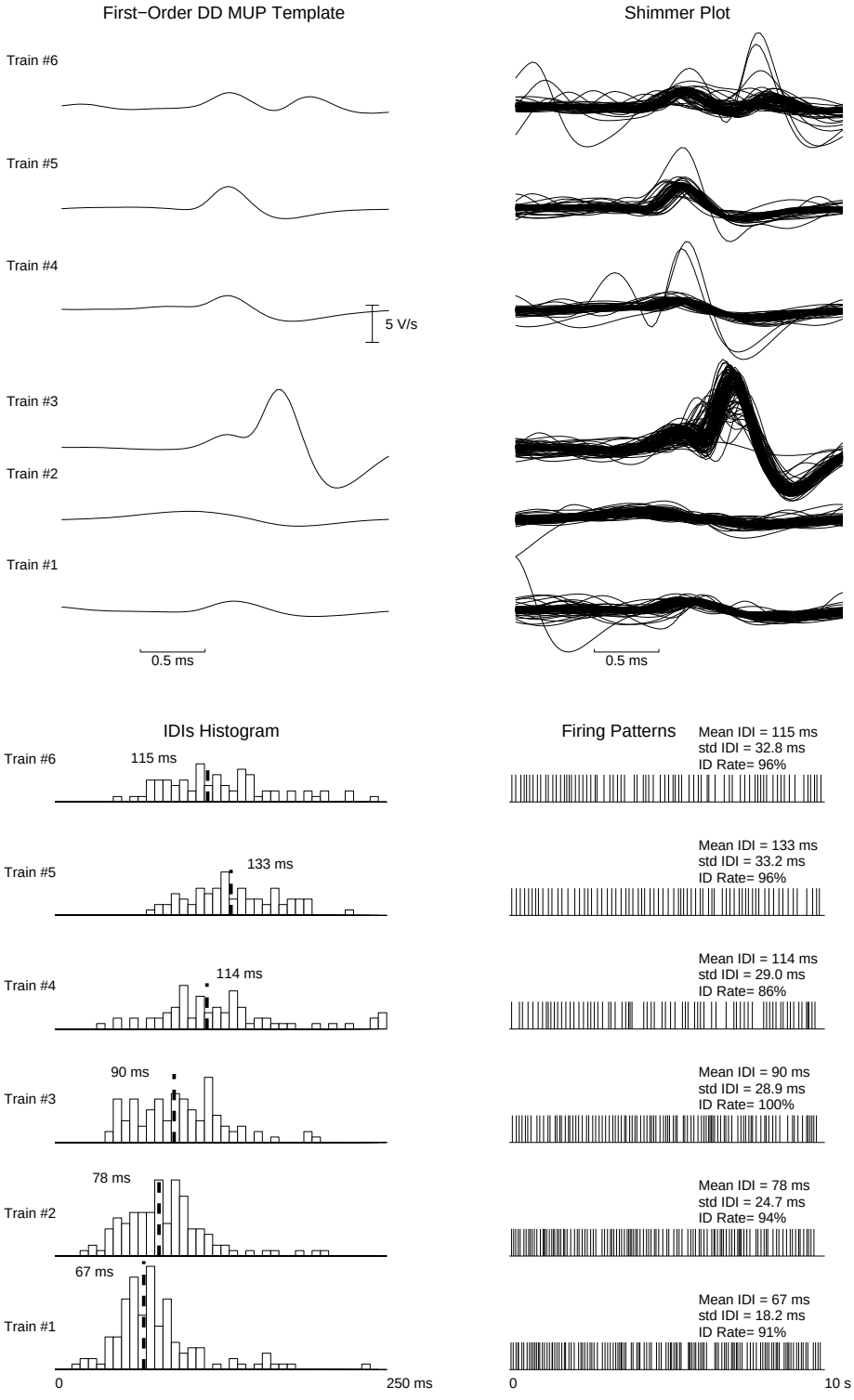

Figure 5.18: Signal 2-4 decomposition using the 6 out of 16 diversity-based ensembles hybrid classifier fusion consisting of majority voting and Sugeno fuzzy integral aggregator.

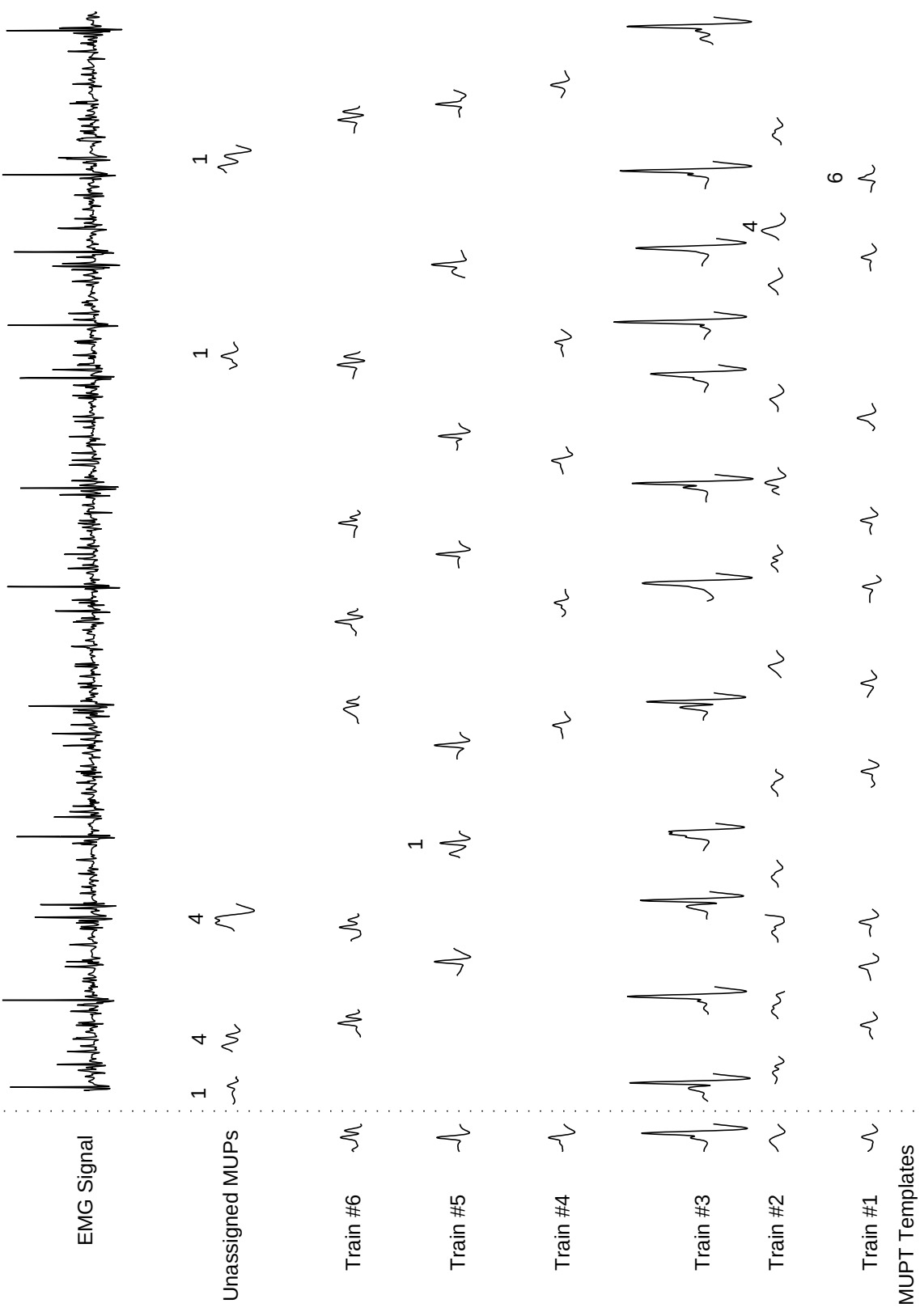

Figure 5.19: MUP traces for a 1 s from 4 s to 5 s interval decomposition results of signal 2-4 using the 6 out of 8 diversity-based adaptive hybrid classifier fusion approach consisting of majority voting and Sugeno fuzzy integral aggregator.

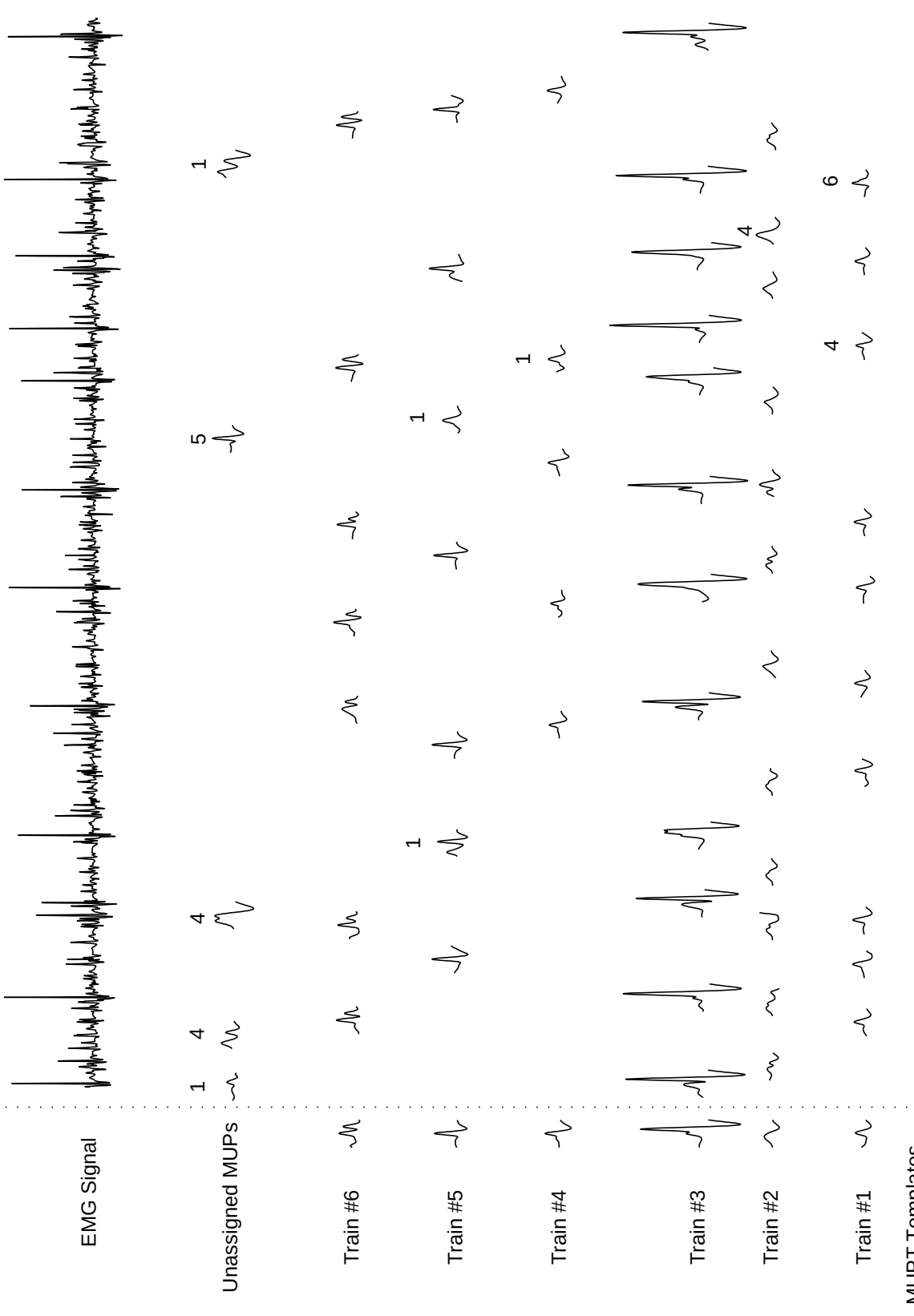

Figure 5.20: MUP traces for a 1 s from 4 s to 5 s interval decomposition results of signal 2-4 using the 6 out of 16 diversity-based adaptive hybrid classifier fusion approach consisting of majority voting and Sugeno fuzzy integral aggregator.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we studied the effectiveness of using classifier fusion of ensembles of base classifiers for EMG signal decomposition with an objective to improve performance and robustness. To achieve this goal, we explored many classification paradigms and adapted them to the problem we are investigating, evaluated the developed classifiers using simulated and real EMG signals of different complexities, refined the misclassification in created MUPTs through proposing a set of IDI statistics capable of detecting erroneous MUP classifications, proposed and tested a new hybrid classifier fusion approach for improving the results, and finally adopted an iterative adaptive MUP classification approach for train-wise adjustment of each MUPT assignment threshold based on train firing pattern statistics to exclude MUPs causing firing pattern inconsistencies.

The classification task in EMG signal decomposition deals with two kinds of data: the MUP shapes and the time of occurrences of MUPs in such a way that the classification of MUPs can not be considered independent of MU firing pattern constraints. Therefore, for the purpose of reducing erroneous MUP classifications to a reasonable extent we devised IDI statistics for the detection of MUPT misclassifications and then adopted the adaptive

setting of a MUPT assignment threshold based on these statistics. The net affect of the adaptive setting of the MUPT assignment threshold over the sets of EMG signals studied was an increase in the number of MUP classifications with only a minor increase in the number of incorrect classifications resulting in overall improved and less variable classification rates.

Each EMG signal is a classification problem in itself and there is no a priori knowledge about the distribution of MUP data, where there are no a priori probabilities and no MUPT conditional densities. Due to the lack of this information, we explored non-parametric classification procedures based on a certainty measure, nearest neighbour rule, and a similarity measure.

We developed and used four different kinds of base classifiers that work on time-domain and wavelet-domain types of features. Base classifiers consist of the Certainty classifiers and its adaptive versions, classifiers based on the nearest neighbour decision rule: the fuzzy k-NN and the adaptive fuzzy k-NN classifiers, and classifiers that use correlation measures: the normalized cross correlation and the pseudo-correlation as an estimation of the degree of similarity between a pattern and a class template: the matched template filter classifiers and its adaptive counterpart.

The system we developed is capable of combining the outputs of the heterogeneous base classifiers in such away that it takes the class labels decided by each base classifier for each MUP or the confidence assignment values generated by each base classifier for each MUP and fuses them using classifier fusion schemes to produce a combined decision with better accuracy.

The developed system capabilities have been extended by incorporating a new hybrid classifier fusion approach in addition to the one level classifier fusion scheme. The hybrid classifier fusion approach has been described and evaluated using simulated and real EMG signals and compared with the performance of the ensemble of constituent base classifiers and also compared with the performance of a one level classifier fusion approach consisting

of either the majority voting, fixed combination rules, or fuzzy integral fusion. In these situations and across the EMG signal data sets, the hybrid approach had better average classification performance overall, especially in terms of the classification errors.

The developed system capabilities have been extended further by proposing a diversity-based hybrid classifier fusion approach to overcome the limitation of selecting the base classifiers comprising the ensemble, where there is a need to perform an exhaustive search for the best accuracy classifier ensemble. The improvement in accuracy of a classifier fusion system does not depend only on the fusion method used for combining the base classifiers but also on the selection of classifiers used for the combination. Based on this, we modified the hybrid classifier fusion approach by exploiting a diversity measure for designing classifier teams. We chose the kappa statistics measure for this purpose to estimate the level of agreement between the base classifier outputs, i.e., to measure the degree of decision similarity between the base classifiers.

When combining base classifiers having the best performance does not mean that the classifier fusion system will give the optimal performance. Also if some of the base classifiers in the ensemble having weak classification performance does not mean that the weak ones when selected to work as a team with others will not improve the performance. This behaviour was noticed when the ensemble contains some matched template filters base classifiers having weak performances relative to the certainty-based and the fuzzy k-NN classifiers.

6.2 Research Contributions

In the author's opinion, the original contribution of the work are:

1. Designed, adapted, and adopted different type classifiers for the MUP classification task in EMG signal decomposition.

2. Formulated a set of IDI statistics for detecting erroneous MUP classifications.

3. Devised an iterative adaptive approach for train-wise adjustment of MUPT assignment threshold based on MU firing pattern statistics.

4. Proposed a hybrid classifier fusion approach that uses an aggregator module consisting of two classifier combiners working in a complementary manner.

5. Proposed a diversity-based hybrid classifier fusion approach that follows the overproduce and choose strategy. The proposed approach allows the automatic selection of classifier ensembles from classifiers pool based on the level of agreement among the classifiers and as determined by the kappa statistics.

6.3 Future Work

For extending this research, we suggest exploring the following ways as a supplement to multiple classifier fusion approach for further and investigation aiming to further improve the performance of the classification task in the EMG signal decomposition process:

1. Using the local discriminant bases (LDB) algorithm [88], [89] to construct an orthonormal basis, i.e., a redundant set of wavelet packet bases having a binary tree structure, which maximizes the class separability for the classification task in the EMG signal decomposition. The LDB algorithm relies on the best basis paradigm [17] through searching for a wavelet packet basis in a dictionary that best illuminates the differences among classes by using some class separability measure such as *relative entropy*. The LDB reduces the dimensionality of the problem by using these basis functions, which are well localized in the time-frequency plane, as feature extractors.

2. The fuzzy densities $g^{i/k}$ play an essential role through the determination of the fuzzy measures used in the fusion process of the fuzzy integral method. Instead of estimating $g^{i/k}$ through the recognition accuracies of each class with each base classifier used

in this thesis, the densities $g^{i/k}$ values may be estimated either using an adaptive fuzzy integral [10], [79] or using genetic algorithms to search for an optimal set of $g^{i/k}$ values [80], [112].

3. Using the fuzzy integral as a supervised classifier. There are two modes of use of the fuzzy integral: the classifier-level mode in which the fuzzy integral is used as an aggregator; and the feature-level mode in which the fuzzy integral is used as a supervised classifier. In this thesis, the fuzzy integral has been used at the classifier-level as a trainable aggregator in the classifier fusion domain. The fuzzy integral in the feature-level is able to model some kind of interaction between features [39]. Each component of the pattern feature vector is considered a data source and it provides a degree of confidence for the assignment of the pattern to a specific class, then these partial degrees of confidence are combined in a consensus-like manner by the fuzzy integral [37], [38], [40], [71].

4. Re-designing the certainty-based classifiers so that they generate confidence values for all respective MUPTs instead of just the most closest and the second closest MUPTs.

5. The adaptive setting of MUPT threshold used in this thesis adopted fixed values for the firing pattern consistency statistics and based on that the MUPT assignment threshold is increased or decreased. Instead, we may train a classifier to check the consistency of a MUPT and based on its decision, the MUPT assignment threshold is adjusted.

6. As stated in Section 6.1, each EMG signal is a classification problem in itself and there is no a priori knowledge about the distribution of MUP data. Therefore, due to the lack of this information and to enable the training for base classifiers and the data-dependent classifier fusion schemes, there is a need to construct a database of different complexities EMG signals, both simulated and real, to extract MUP

features that make the learning of classification approaches practical to a certain extent. This will enable base classifiers combined in a classifier fusion system to be evaluated using the MUP features stored in the database and then we use a weighting system to assign worth values of how important each base classifiers is relative to the MUP features.

6.4 Publications

1. S. Rasheed, D. Stashuk, and M. Kamel, *A classifier fusion modelling environment for motor unit potential classification*, to be submitted to ACM Transactions on Modeling and Computer Simulation Journal.

2. S. Rasheed, D. Stashuk, and M. Kamel, *An interactive modelling environment for motor unit potential sorting using certainty-based classifiers*, to be submitted to Simulation Modelling Practice and Theory Journal.

3. S. Rasheed, D. Stashuk, and M. Kamel, *Integrating heterogeneous classifier ensembles for EMG signal decomposition classification task based on classifier agreement*, to be submitted to IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans Journal.

4. S. Rasheed, D. Stashuk, and M. Kamel, *Multiclassifier fusion for motor unit potential sorting*, Submitted for publication in Pattern Recognition Journal.

5. S. Rasheed, D. Stashuk, and M. Kamel, *A hybrid classifier fusion approach for motor unit potential classification during EMG signal decomposition*, Submitted for publication in IEEE Transactions on Biomedical Engineering Journal.

6. S. Rasheed, D. Stashuk, and M. Kamel, *Adaptive fuzzy k-NN classifier for EMG signal decomposition*, Accepted for publication in Medical Engineering & Physics Journal and now is published on-line at www.sciencedirect.com.

7. S. Rasheed, D. Stashuk, and M. Kamel, *Adaptive certainty-based classification for decomposition of EMG signals*, Medical & Biological Engineering & Computing, Vol. 44, No. 4, pp. 298-310, 2006.

8. S. Rasheed, D. Stashuk and M. Kamel, *Multi-classification techniques applied to EMG signal decomposition*, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, SMC 04, The Hugue, The Netherland, Vol. 2, pp. 1226-1231, 2004.

Bibliography

[1] L. A. Alexandre, A. C. Campilho, and M. Kamel. On combining classifiers using sum and product rules. *Pattern Recognition Letters*, 22:1283–1289, 2001.

[2] H. Altinçay and M. Demirekler. Undesirabel effects of normalization in multiple classifier systems. *Pattern Recognition Letters*, 24:1163–1170, 2003.

[3] S. Andreassean. Methods for computer-aided measurement of motor unit parameters. *The London Symposia - Supplement 39 to Electroencephalography and Clinical Neurophysiology, 13-20*, pages 13–20, 1987.

[4] A. F. Atiya. Recognition of multiunit neural signals. *IEEE Transactions on Biomedical Engineering*, 39(7):723–729, 1992.

[5] J. V. Basmajian and C. J. De Luca. *Muscles Alive: Their Functions Revealed by Electromyography*. Williams and Wilkins, Baltimore MD, 5th edition, 1985.

[6] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, Cambridge, MA, 1975.

[7] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[8] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 17(10):955–966, 1995.

[9] E. Chauvet, O. Fokapu, J.-Y. Hogrel, D. Gamet, and J. Duchêne. Automatic identification of motor unit action potential trains from electromyographic signals using fuzzy techniques. *Medical & Biological Engineering & Computing*, 41:646–653, 2003.

[10] Z. Chi, H. Yan, and T. Pham. World Scientific, 1996.

[11] Sung-Bae Cho. Neural-network classifiers for recognizing totally unconstrained handwritten numerals. *IEEE Transactions on Neural Networks*, 8(1):43–53, 1997.

[12] Sung-Bae Cho and J. H. Kim. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Transactions on Systems, Man and Cybernetics*, 25(2):380–384, 1995.

[13] Sung-Bae Cho and J. H. Kim. Multiple network fusion using fuzzy logic. *IEEE Transactions on Neural Networks*, 6(2):497–501, 1995.

[14] C. I. Christodoulou and C. S. Pattichis. Medical diagnostic systems using ensembles of neural SOFM classifiers. In *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems ICECS '99*, volume 1, pages 121–124, Turkey, 1999.

[15] C. I. Christodoulou, C. S. Pattichis, and W. F. Fincham. A modular neural network decision support system in EMG diagnosis. *Journal of Intelligent Systems*, 8(1):99–143, 1998.

[16] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(2):37–46, 1960.

[17] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithm for best basis selection. *IEEE Transactions on Information Theory*, 38:713–719, 1992.

[18] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

[19] T. G. Dietterich. Machine learning research: four current directions. *AI Magazine*, 18(4):97–136, 1997.

[20] D. L. Donoho. Wavelet shrinkage and w.v.d.: A 10-minute tour. Technical report, Standford University, CA, 1992.

[21] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transaction on Information Theory*, 41(3):613–627, 1995.

[22] R. O. Duda, P.E Hart, and D. G. Stork. *Pattern Classification.* John Wiley & Sons, 2 edition, 2001.

[23] R. P. W. Duin. The combining classifier: to train or not to train? In *Proceedings of the 16th International Conference on Pattern Recognition*, volume 2, pages 765–770, 2002.

[24] R. P. W. Duin and D. M. J. Taz. Experiments with classifier combining rules. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, Lecture Notes in Computer Science, 1857, pages 16–29, Gagliari, Italy, 2000. Springer.

[25] H. Etawil and D. Stashuk. Resolving superimposed motor unit action potentials. *Medical & Biological Engineering & Computing*, 34(1):33–40, 1996.

[26] H. A. Y. Etawil. Motor unit action potentials: discovering temporal relations of their trains and resolving their superpositions. Master's thesis, University of Waterloo., Waterloo, Ontario, Canada, 1994.

[27] J. Fang, G. C. Agarwal, and B. T. Shahani. Decomposition of EMG signal by wavelet spectrum matching. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1253–1256, 1997.

[28] J. Fang, G. C. Agarwal, and B. T. Shahani. Decomposition of multiunit electromyogram signals. *IEEE Transactions on Biomedical Engineering*, 46(6):685–697., 1999.

[29] D. Farina, R. A. Colombo, R. Merletti, and H. B. Olsen. Evaluation of intra-muscular EMG signal decomposition algorithms. *Journal of Electromyography and Kinesiology*, 11:175–187, 2001.

[30] D. Farina, A. Crosetti, and R. Merletti. A model for the generation of synthetic intramuscular EMG signals to test decomposition algorithms. *IEEE Transactions on Biomedical Engineering*, 48(1):66–77, 2001.

[31] J. L. Fleiss, B. Levin, and M. C. Paik. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 3 edition, 2003.

[32] J. R. Florestal, P. A. Mathieu, and A. Malanda. Automatic decomposition of simulated EMG signals. In *Proceedings of the 28th Conference of the Canadian Medican and Biological Engineering Society*, pages 29–30, 2004.

[33] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. In *Proceedings of the Thirtenth International Conference on Machine Learning*, pages 149–156, 1996.

[34] G.-H.Chiang. Choquet fuzzy integral-based hierarchical networks for decision analysis. *IEEE Transactions on Fuzzy Systems*, 7(1):63–71, 1999.

[35] A. B. Geva. Application of fuzzy clustering to biomedical signal processing and dynamic system identification. In A. Metin, editor, *Nonlinear Biomedical Signal Processing*, volume 1, pages 27–52. IEEE Press, New York, NY, 2000.

[36] G. Giacinto and F. Roli. An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters*, 22:25–33, 2001.

[37] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89:445–456, 1996.

[38] M. Grabisch. The representation of importance and interaction of features by fuzzy measures. *Pattern Recognition Letters*, 17:567–575, 1996.

[39] M. Grabisch. Fuzzy integral for classification and feature extraction. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals*, pages 415–434. North-Holland, 2000.

[40] M. Grabisch and J.-M. Nicolas. Classification by fuzzy integrals: performace and tests. *Fuzzy Sets and Systems*, 65:255–271, 1994.

[41] R. Gut and G. S. Moschytz. High-precision EMG signal decomposition using communication techniques. *IEEE Transactions on Signal Processing*, 48(9):2487–2494, 2000.

[42] A. Hamilton-Wright and D. W. Stashuk. Physiologically based simulation of clinical EMG signals. *IEEE Transactions on Biomedical Engineering*, 52(2):171–183, 2005.

[43] S. Hashem. Optimal linear combinations of neural networks. *Neural Networks*, 10(4):599–614, 1997.

[44] M. H. Hassoun, C. Wang, and R. Spitzer. Nnerve: Neural network extraction of repetitive vectors for electromyography - part i: algorithm. *IEEE Transactions on Biomedical Engineering*, 41(11):1039–1052, 1994.

[45] M. H. Hassoun, C. Wang, and R. Spitzer. Nnerve: Neural network extraction of repetitive vectors for electromyography - part ii: performance analysis. *IEEE Transactions on Biomedical Engineering*, 41(11):1053–1061, 1994.

[46] T. K. Ho, J. J. Hull, and S. N. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 16(1):66–75.

[47] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[48] A. K. Jain and M. D. Ramaswami. Classifier design with parzen windows. In E. S. Gelsema and L. N. Kanal, editors, *Pattern Recognition and Artificial Intelligence*, pages 211–228. Elsevier Science Publishers B. V., 1988.

[49] A. Jóźwik. A learning scheme for a fuzzy k-nn rule. *Pattern Recognition Letters*, 1:287–289, 1983.

[50] M. S. Kamel and N. M. Wanas. Data dependence in combining classifiers. In T. Windeatt and F. Roli, editors, *Multiple Classifier Systems*, Lecture Notes in Computer Science, 2709, pages 1–14, Guilford, UK, 2003. Springer.

[51] S. Kanoun. A new digital signal processing technique for the estimation of motor unit action potential templates. Master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 1999.

[52] J. M. Keller, P. Gader, H. Tahani, J.-H. Chiang, and M. Mohamed. Advances in fuzzy integration for pattern recognition. *Fuzzy Sets and Systems*, 65:273–283, 1994.

[53] J. M. Keller, M. R. Gray, and J. A. Givens JR. A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Systems, Man, and Cybernetics*, 15(4):580–585, July/August 1985.

[54] J. M. Keller and J. Osborn. Advances in fuzzy integration for pattern recognition. *International Journal of Approximate Reasoning*, 15(1):1–24, 1996.

[55] J. M. Kittler, M. Hatef, R. P. V. Duin, and J. Matas. On combining classifiers. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 20(3):955–966, 1998.

[56] L. I. Kuncheva. *Fuzzy Classifier Design*. Studies in Fuzziness and Soft Computing. Physica-Verlag, 2000.

[57] L. I. Kuncheva. Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34:299–314, 2001.

[58] L. I. Kuncheva and C. J. Whitacker. Measure of divesity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51:181–207, 2003.

[59] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin. Is independence good for combining classifiers? In *Proceedings of the 15th International Conference on Pattern Recognition*, volume 2, pages 168–171, 2001.

[60] P. Lago and N. B. Jones. Effect of motor-unit firing time statistics on e.m.g. spectra. *Medical & Biological Engineering & Computing*, 15(11):648–655, 1977.

[61] L. Lam. Classifier combinations: implementations and theoretical issues. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, Lecture Notes in Computer Science, 1857, pages 77–86. Springer, Cagliari, Italy, 2000.

[62] L. Lam and C. Y. Suen. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transaction on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 27(5):553–568, 1997.

[63] R. S. LeFever and C. J. De Luca. A procedure for decomposing the myoelectric signal into its constituent action potentials - part i: technique, theory, and implementation. *IEEE Transactions on Biomedical Engineering*, 29(3):149–157, 1982.

[64] G. H. Loudon, N. B. Jones, and A. S. Sehmi. New signal processing techniques for the decompositon of emg signals. *Medical & Biological Engineering & Computing*, 30(11):591–599, 1992.

[65] C. J. De Luca. Physiology and mathematics of myoelectric signals. *IEEE Transactions on Biomedical Engineering*, 26(6):313–325, 1979.

[66] C. J. De Luca. Reflections on EMG signal decomposition. In J. E. Desmedt, editor, *Computer-Aided Electromyography and Expert Systems*, pages 33–37. Elsevier Science, Amesterdam, The Netherlands, 1989.

[67] P. B. C. Matthews. Relationship pf firing intervals of human motor units to the trajectory of post-spike after-hyperpolarization and synaptic noise. *Journal of physiology*, 492(2):597–628, 1996.

[68] K. C. McGill. *A method for quantitating the clinical electromyogram.* Ph.d. dissertation, Stanford University, Stanford, CA, 1984.

[69] K. C. McGill, K. Cummins, and L. J. Dorfman. Automatic decomposition of the clinical electromyogram. *IEEE Transactions on Biomedical Engineering*, 32(7):470–477, 1985.

[70] S. Medasani, J. Kim, and R. Krishnapuram. An overview of membership function generation techniques for pattern recognition. *International Journal of approximate reasoning*, 19:391–417, 1998.

[71] L. Mikenina and H.-J. Zimmermann. Improved feature selectin and classification by the 2-additive fuzzy measure. *Fuzzy Sets and Systems*, 107:197–218, 1999.

[72] K. Mirfakhraei and K. Horch. Recognition of temporally changing action potentials in multiunit neural recordings. *IEEE Transactions on Biomedical Engineering*, 44(2):123–131, 1997.

[73] S. Hamid Nawab, R. Wotiz, and C. J. De Luca. Improved resolution of pulse superpositions in a knowledge-based system EMG decomposition. In *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 69–71, 2004.

[74] M. Nikolic, J. A. Sørensen, K. Dahl, and C. Krarup. Detailed analysis of motor unit activity. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society Conference*, pages 1257–1260, 1997.

[75] G. M. Paoli. Estimating certainty in classification of motor unit action potentials. Master's thesis, University of Waterloo., Waterloo, Ontario, Canada, 1993.

[76] D. Partridge and W. B. Yates. Engineering multiversion neural-net systems. *Neural Computation*, 8:869–893, 1996.

[77] D. Perkel, G. L. Gerstein, and G. P. Moore. Neural spike trains and stochastic point process - part i. the single spike train. *The Biophysical Journal*, 7:391–418, 1967.

[78] M. P. Perrone. Averaging/modular techniques for neural networks. In M. A. Arbib, editor, *The handbook of brain theory and neural networks*, pages 126–129. MIT Press, Cambridge, Massachusetts, 1995.

[79] T. D. Pham. Combination of multiple classifiers using adaptive fuzzy integral. In *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems ICECS '99*, pages 50–55, 2002.

[80] T. D. Pham and H. Yan. Fusion of handwritten numeral classifiers based on fuzzy and genetic algorithms. In *Proceedings of the 1997 Annual Meeting of the North American Fuzzy Information Processing Society NAFIPS '97*, pages 257–262, 1997.

[81] Y. Qu. Automatic clustering of motor unit action potentials using shape and time information. Master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 1993.

[82] A. F. R. Rahman and M. C. Fairhurst. An evaluation of multi-expert configurations for the recognition of handwritten numerals. *Pattern Recognition*, 31(9):1225–1273, 1998.

[83] S. Rasheed, D. Stashuk, and M. Kamel. Adaptive fuzzy k-nn classifier for EMG signal decomposition. *accepted for publication in the Medical Engineering & Physics journal and now is published on line at www.sciencedirect.com.*

[84] S. Rasheed, D. Stashuk, and M. Kamel. Multi-classification techniques applied to EMG signal decomposition. In *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics, SMC 04*, volume 2, pages 1226–1231, The Hugue, The Netherland, 2004.

[85] S. Rasheed, D. Stashuk, and M. Kamel. Adaptive certainty-based classification for decomposition of EMG signals. *Medical & Biological Engineering & Computing*, 44(4):298–310, 2006.

[86] F. Roli and G. Giacinto. Design of multiple classifier systems. In H. Bunke and A. Kandel, editors, *Hybrid methods in pattern recognition*, pages 199–226. World Scientific, Cambridge, Massachusetts, 2002.

[87] F. Roli, G. Giacinto, and G. Vernazza. Methods for designing multiple classifier systems. In J. Kittler and F. Roli, editors, *Proceedings of the Second International Workshop on Multiple Classifier Systems*, Lecture Notes in Computer Science, 2096, pages 78–87. Springer-Verlag, 2001.

[88] N. Saito. *Local feature extraction and its applications using a library of bases.* Ph.d. dissertation, Yale University, New Haven, CT, 1994.

[89] N. Saito and R. R. Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5:337–358, 1995.

[90] R. E. Schapire. The strength of weal learnability. *Machine Learning*, 5(2):197–227, 1990.

[91] J. Schürmann. *Pattern Classification: A Unified View of Statistical and Neural Approaches*. John Wiley & Sons, New York, 1996.

[92] A. J. C. Sharkey, N. E. Sharkey, U. Gerecke, and G. O. Chandroth. The ẗest and selectäpproach to ensemble combination. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, Lecture Notes in Computer Science, 1857, pages 30–44, Gagliari, Italy, 2000. Springer.

[93] C. A. Shipp and L. I. Kuncheva. Relationships between combination methods and measures of diversity in combining classifiers. *Information Fusion*, 3:135–148, 2002.

[94] R. Spitzer and M. H. Hassoun. Nnerve: neural network extraction of repetitive vectors for electrodiagnosis. In *Proceedings of Sixth Annual IEEE Symposium on Computer-Based Medical Systems*, pages 171–176, 1993.

[95] E. Stålberg, B. Flack, M. Sonoo, S. Stålberg, and M. Åström. Multi-mup EMG analysis - a two year experience in daily clinical work. *Electroencephalography and Clinical Neurophysiology*, 97:145–154, 1995.

[96] D. W. Stashuk. Decomposition and quantitative analysis of clinical electromyographic signals. *Medical Engineering & Physics*, 21:389–404, 1999.

[97] D. W. Stashuk. EMG signal decomposition: how can it be accomplished and used? *Journal of Electromyography and Kinesiology*, 11:151–173, 2001.

[98] D. W. Stashuk and H. de Bruin. Automatic decomposition of selective needle-detected myoelectric signals. *IEEE Transactions on Biomedical Engineering*, 35(1):1–10, 1988.

[99] D. W. Stashuk and G. M. Paoli. Robust supervised classification of motor unit action potentials. *Medical & Biological Engineering & Computing*, 36(1):75–82, 1998.

[100] D. W. Stashuk and Y. Qu. Adaptive motor unit action potential clustering using shape and time information. *Medical & Biological Engineering & Computing*, 34(1):41–49, 1996.

[101] D. W. Stashuk and Y. Qu. Robust method for estimating motor unit firing-pattern statistics. *Medical & Biological Engineering & Computing*, 34(1):50–57, 1996.

[102] L. D. Stefano and S. Mattoccia. Fast template matching using bounded partial correlation. *Machine Vision and Applications*, 13:213–221, 2003.

[103] C. Suen and L. Lam. Multiple classifier combination methodologies for different output levels. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, Lecture Notes in Computer Science, 1857, pages 52–66, Gagliari, Italy, 2000. Springer.

[104] M. Sugeno. *Theory of fuzzy integrals and its applications*. Ph.d. dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[105] M. Sugeno. Fuzzy measures and fuzzy integrals - a survey. In *Fuzzy Automata and Decision Processes*, pages 89–102. North-Holland, Amsterdam, 1977.

[106] H. Tahani and J. M. Keller. Information fusion in computer vision using the fuzzy integral. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(3):733–741, 1990.

[107] S. Usaui and I. Amidror. Digital low-pass differentiation for biological signal processing. *IEEE Transactions on Biomedical Engineering*, 29(10):686–693, 1982.

[108] G. Valentini. *Ensemble methods based on biasvariance analysis*. Ph.d. dissertation, University of Genova, Genova, Italy, 2003.

[109] G. Valentini and F. Masulli. Ensembles of learning machines. In M. Marinaro and R. Tagliaferri, editors, *Neural Nets: 13th ItalianWorkshop on Neural Nets*, Lecture Notes in Computer Science, 2486, pages 3–20, Vietri sul Mare, Italy, 2002. Springer-Verlag.

[110] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis. Soft combination of neural classifiers: A comparative study. *Pattern Recognition Letters*, 20:429–444, 1999.

[111] N. Wanas. *Feature based architecture for decision fusion.* Ph.d. dissertation, University of Waterloo, Waterloo, Ontario, Canada, 2003.

[112] D. Wang, X. Wang, and J. M. Keller. Determining fuzzy integral densities using a genetic algorithm for pattern recognition. In *Proceedings of the 1997 Annual Meeting of the North American Fuzzy Information Processing Society NAFIPS '97*, pages 21–24, 1997.

[113] D. H. Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 1992.

[114] K. Woods. Combination of multiple classifiers using local accuracy estimates. *IEEE Transaction in Pattern Analysis and Machine Intelligence*, 4(4):405–410, 1997.

[115] L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transaction on Systems, Man, and Cybernetics*, 22(3):418–435, 1992.

[116] D. Zennaro, P. Wellig, V. M. Koch, G. S. Moschytz, and T. Läubli. A software package for the decomposition of long-term multi-channel EMG signals using wavelet coefficients. *IEEE Transactions on Biomedical Engineering*, 50(1):58–69, 2003.

[117] D. Zennaro, P. Wellig, G. S. Moschytz, T. Läubli, and H. Krueger. A decomposition software package for the decomposition of long-term multi-channel electromyographic signals. In *Proceedings of the 23th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1070–1073, Turkey, 2001.