

Classification of Hand Gestures Using Time-Frequency Analysis and Different Artificial Intelligence Methods

Submitted to the Graduate School of Natural and Applied Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

in Biomedical Engineering

by

Deniz Hande K1sa

ORCID 0000-0002-5882-0605

July, 2023

This is to certify that we have read the thesis **Classification of Hand Gestures Using Time-Frequency Analysis and Different Artificial Intelligence Methods** submitted by **Deniz Hande Kısa**, and it has been judged to be successful, in scope and in quality, at the defense exam and accepted by our jury as a MASTER'S THESIS.

APPROVED BY:

Advisor: **Asst. Prof. Dr. Onan GÜREN**
İzmir Kâtip Çelebi University

Co-advisor: **Prof. Dr. Aydın AKAN**
İzmir University of Economics

Committee Members:

Asst. Prof. Dr. Ebru SAYILGAN
İzmir University of Economics

Asst. Prof. Dr. İbrahim KAYA
İzmir Kâtip Çelebi University

Date of Defense: July 25, 2023

Declaration of Authorship

I, **Deniz Hande Kİsa**, declare that this thesis titled **Classification of Hand Gestures Using Time-Frequency Analysis and Different Artificial Intelligence Methods** and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for the Master's degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. This thesis is entirely my own work, with the exception of such quotations.
- I have acknowledged all major sources of assistance.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Date: 03.07.2023

Classification of Hand Gestures Using Time-Frequency Analysis and Different Artificial Intelligence Methods

Abstract

Hand gesture-based systems are one of the most effective technological advances. Surface electromyography (sEMG) is utilized as a popular input data for gesture classification with elevated accuracy and advanced control capability. In this thesis, which is based on the classification performance of Hilbert-Huang spectrum (HHS) images obtained from Hilbert Huang Transform (HHT) of the sEMG of the gestures, an evaluation of the results of artificial intelligence (AI) methods on hand gesture classification using HHS image is presented. A dataset of 4-channel sEMG of seven hand movements was used. HHS images were obtained by applying HHT to sEMG signals. In hand gesture classification, six image features obtained from the gray-level co-occurrence matrix (GLCM) of HHS images were classified in machine learning (ML) and fuzzy logic (FL) models. The GLCM features were also extracted from the EMG and intrinsic mode functions (IMF) snapshots and classified in order to make comparisons within ML. In FL, the GLCM features were classified using two different clustering techniques, subtractive clustering (SC) and fuzzy c-mean (FCM) clustering. In addition, HHS images are used directly in deep learning (DL) models and classified into seven different convolutional neural networks (CNN) architectures based on the transfer learning method. Also, the effects of different IMF combinations and different signal lengths on classification performance were evaluated while generating HHS images. Three different AI methods were evaluated within themselves. According to

the results obtained within the scope of the study, it was determined that in all three AI methods, HHS images and features based on time-frequency analysis were successful in hand gesture classification and each of them could be used as an alternative method.

Keywords: Hand gestures, time-frequency analysis, machine learning, deep learning, fuzzy logic

Zaman-Frekans Analizi ve Farklı Yapay Zeka Yöntemleri Kullanılarak El Hareketlerinin Sınıflandırılması

ÖZ

El hareketi tabanlı sistemler, en etkili teknolojik gelişmelerden biridir. Yüzeysel elektromiyografisi (sEMG), yüksek doğruluk ve kontrol yeteneği ile hareket sınıflandırması için popüler bir veri olarak kullanılır. Hareketlerin sEMG'sinin Hilbert-Huang Dönüşümü'nden (HHT) elde edilen Hilbert-Huang spektrumu (HHS) görüntülerinin sınıflandırma performansına dayanan bu tezde, yapay zeka (AI) yöntemlerinin HHS görüntüsü kullanarak el hareketi sınıflandırmasına ilişkin sonuçların değerlendirilmesi sunulmaktadır. Yedi el hareketine ait 4 kanallı sEMG veri seti kullanıldı. sEMG sinyallerine HHT uygulanıp HHS imgeleri elde edildi. El hareketi sınıflandırmada HHS görüntülerinin gri düzey eş oluşturma matrisinden (GLCM) elde edilen altı imge özneliği makine öğrenmesi (ML) ve bulanık mantık (FL) modellerinde sınıflandırılmıştır. ML'de kendi içinde karşılaştırma yapmak için GLCM öznelikleri, EMG ve içsel mod fonksiyonu (IMF) ekran görüntülerinden çıkarılıp sınıflandırılmıştır. FL'de ise GLCM öznelikleri azaltımlı ve bulanık c-ortalama kümeleme teknikleri kullanılarak ayrı ayrı sınıflandırılmıştır. Bunlarla birlikte, derin öğrenme (DL) modellerinde HHS imgeleri doğrudan kullanılarak transfer öğrenimi yöntemine dayalı yedi farklı DL mimarisinde sınıflandırılmıştır. Burada ayrıca, HHS imgeleri oluşturulurken farklı IMF kombinasyonlarının ve farklı

sinyal uzunluklarının sınıflandırma performansına etkisi değerlendirilmiştir. Üç farklı AI yönteminin kendi içinde değerlendirilmesi yapılmıştır. Çalışma kapsamında elde edilen üç AI yönteminin sonuçlara göre, zaman-frekans analizine dayalı HHS imge ve özniteliklerinin el hareketi sınıflandırmada başarılı olduğu ve her bir yaklaşımın alternatif bir yöntem olarak kullanılabileceği belirlenmiştir.

Anahtar Kelimeler: El hareketleri, zaman-frekans analizi, makine öğrenmesi, derin öğrenme, bulanık mantık

To my mother...

Acknowledgment

I would like to sincerely thank my supervisors Asst. Prof. Dr. Onan Güren and Prof. Dr. Aydın Akan for their support and valuable knowledge. I would also like to express my sincere thanks and gratitude to my supervisor Asst. Prof. Dr. Onan Güren for her guidance in all matters, and for giving me the opportunity to work in her laboratory and projects.

I would like to acknowledge The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2210-C National MSc/MA Scholarship Program in the Priority Fields in Science and Technology for financial support of this study under grant 1649B022103607. Also, Izmir Katip Celebi University Scientific Research Projects Coordination Unit (grant number 2022-GAP-MÜMF-0001).

I would like to especially thank Res. Asst. Mehmet Akif Özdemir and Res. Asst. Gizem Dilara Özdemir for their assistance, guidance, and contributions during the whole process of my M.Sc.

Also, I would like to thank my mother Nuray Akgül, my parents, and my family for supporting and encouraging me during this period of my M.Sc.

July - 2023

Deniz Hande K1sa

Table of Contents

Declaration of Authorship.....	ii
Abstract.....	iii
Öz.....	v
Acknowledgment	viii
Table of Contents	ix
List of Figures	xii
List of Tables.....	xvi
List of Abbreviations.....	xvii
List of Symbols	xxi
1 Introduction.....	1
1.1 Hand Gesture Recognition.....	1
1.2 sEMG Signal and Analysis	3
1.2.1 sEMG Generation.....	4
1.2.2 Mathematical Representation of sEMG Signal.....	4
1.2.3 sEMG Signal Preprocessing and Analysis.....	5
1.2.3.1 sEMG Preprocessing.....	6
1.2.3.2 sEMG Signal Analysis.....	6
1.3 Artificial Intelligence Approaches.....	9
1.4 Hand Gesture Classification Studies in the Literature.....	10

1.4.1	Machine Learning-based Studies	11
1.4.2	Deep Learning-based Studies.....	12
1.4.3	Fuzzy Logic-based Studies	13
1.5	Aim of the Study.....	15
2	Materials and Methods.....	16
2.1	sEMG Data.....	17
2.2	Pre-processing.....	18
2.2.1	Filtering.....	19
2.2.2	Segmentation.....	19
2.3	Time-Frequency Analysis.....	21
2.3.1	Hilbert Huang Transform.....	22
2.3.1.1	Empirical Mode Decomposition	22
2.3.1.2	IMF Selection.....	24
2.3.1.3	Hilbert Transform	25
2.3.2	TF Image Fusion and Snapshot Image Generation	26
2.4	Feature Extraction for ML and FL.....	27
2.5	Classification in AI models.....	31
2.5.1	Machine Learning	31
2.5.1.1	Feature Selection.....	31
2.5.1.2	ML Models.....	32
2.5.2	Fuzzy Logic.....	37
2.5.2.1	Adaptive Neuro-Fuzzy Inference System.....	37
2.5.2.2	Clustering Techniques	39
2.5.3	Deep Learning.....	43
2.5.3.1	Architecture of CNN.....	43
2.5.3.2	Transfer Learning.....	48
2.6	Performance Evaluation.....	53
2.7	Statistical Analysis.....	55
3	Results	56
3.1	Results of preporcessing of sEMG Signal	56

3.2	Image Generation.....	57
3.2.1	Results of IMF Selection.....	58
3.2.2	Results of Different Length HHS Images	60
3.2.3	Results of TF Image Fusion	60
3.2.4	Results of HHS Images of Different Gestures	62
3.2.5	Results of Snapshot Images	63
3.3	Feature Matrix Generation	63
3.4	Results of ML-based Classification	64
3.5	Results of FL-based Classification	66
3.6	Results of DL-based Classifications	70
3.6.1	Classification Results of IMF and WL Combination.....	71
3.6.2	Classification Results of CNN Models	73
4	Discussion.....	74
4.1	Evaluation of ML-based Classification Results.....	74
4.1.1	Comparison with State-of-the-Art Studies.....	75
4.2	Evaluation of FL-based Classification Results	76
4.2.1	Comparison with State-of-the-Art Studies.....	77
4.3	Evaluation of DL-based Classification Results	80
4.3.1	Evaluation of IMF and WL Results	80
4.3.2	Evaluation of Results of CNN Models.....	81
4.3.3	Gesture-based Evaluation.....	82
4.3.4	Comparison with State-of-the-Art Studies.....	83
5	Conclusion	89
	References	92
	Appendix	109
	Curriculum Vitae	110

List of Figures

Figure 1.1: Examples of gesture-based systems: (a) EEG-based hand gesture recognition system [4], (b) vision-based hand gesture recognition system [5].....	2
Figure 1.2: A biological signal-based hand movement recognition system: (a) prosthesis control using different biological signals, (b) controlling algorithm for pattern recognition, (c) data collection from healthy muscles of amputee [8].....	2
Figure 1.3: EMG applications areas [11]	3
Figure 1.4: sEMG generation process [14]	4
Figure 1.5: Structure of the recorded sEMG: $g(t)$ is the recorded sEMG signal on which analysis is performed, $x(t)$ is the interested true signal, $e(t)$ is noise, and $H(f)$ is the transfer function of the record tool [15]	5
Figure 1.6: An example of a raw sEMG with 50 Hz interference, the frequency response of a 50 Hz notch filter and a 20–500 Hz band-pass filter, and the resulting filtered signal, respectively [14].....	6
Figure 1.7: The analysis of non-stationary signal using WT: (a) in TD, (b) in FD, and (c) a time-frequency map in TFD [21].....	8
Figure 1.8: The schema of artificial intelligence methods based on this study	9
Figure 1.9: An example schema of EMG-based hand gesture recognition system [31]	11
Figure 2.1: The schema of the method.....	16

Figure 2. 2: Seven hand gestures: (a) Rest position, (b) Extension of wrist, (c) Flexion of wrist, (d) Ulnar deviation of wrist, (e) Radial deviation of wrist, (f) punch (grip), (g) open hand (abduction).....	17
Figure 2.3: Location of electrodes for posterior and anterior views of right arm: (a) 4-channel and ground electrodes placement, and (b) muscles corresponding to electrode positions	18
Figure 2.4: Timeline of data collection procedure	18
Figure 2.5: Overlapping window process for steady-state EMG signal (s: shift length, w: window length)	20
Figure 2.6: Overlapping window process: (a) 4 s steady-state EMG signal, and (b) 200 ms EMG segment.....	21
Figure 2.7: Implementation of EMD to the EMG signal, (a) 200 ms EMG signal, and (b) the visualization of obtained first six IMFs.....	24
Figure 2.8: Statistically significant IMFs selection: (a) the first six extracted IMFs, and (b) the selected first three IMFs	25
Figure 2.9: HHS generation: (a) the selected three IMFs, and (b) Hilbert-Huang Spectrum after application of Hilbert Transform.....	26
Figure 2.10: The visualization of GLCM process.....	29
Figure 2.11: An example visualization of classification of data using SVM.....	34
Figure 2.12: kNN algorithm: the distance between the selected point (yellow) and the neighbor points is calculated with the Euclidean distance formula, and the class is determined according to the proximity and abundance of the neighbors	35
Figure 2.13: An example of ANN structure with three hidden layers	36
Figure 2.14: The structure of ANFIS model	38
Figure 2.15: An example subtractive clustering result of data.....	41

Figure 2.16: An example output of FCM clustering [96]	43
Figure 2.17: An example CNN architecture for image classification [97]	44
Figure 2.18: Convolution operation	45
Figure 2.19: Max Pooling and Sum Pooling operations	46
Figure 2.20: An example of application of activation function, ReLU, to matrix.....	47
Figure 2.21: An example of medical image classification using TL from a study [103]	49
Figure 2.22: Residual blocks of ResNet-34 and ResNet-50	51
Figure 2.23: The seven CNN architectures used in this study	52
Figure 3.1: The plotting of 4-channel filtered sEMG signal of seven hand gestures	56
Figure 3.2: 250 ms-long multi-channel sEMG segments of (a) Rest state, and (b) Extension gesture	57
Figure 3.3: The whole process of obtaining HHS images using first three IMFs of EMG signal: (a) 200 ms sEMG signal, (b) first six IMFs, (c) selected first three IMFs, and (d) HHS image using first three IMFs.....	59
Figure 3.4: HHS images generated using (a) IMF_1 and (b) IMF_3 with 250 ms-long segments.....	60
Figure 3.5: HHS images of the extension movement using first three IMFs of (a) 200 ms, and (b) 250 ms segments	60
Figure 3.6: Fusion image of four channel	61
Figure 3.7: 3D HHS representations: (a) 250 ms-long sEMG segments of Rest state and (b) 250 ms-long sEMG segments of Extension gesture using the first-three IMFs. (c) 2D color-mapped images of corresponding HHS of Rest, and (d) 2D color-mapped images of corresponding HHS of Extension.....	62
Figure 3.8: Snapshot image examples: (a) EMG, and (b) IMFs	63

Figure 3.9: Confusion Matrix of HHS-based classification for ML	65
Figure 3.10: An example of ANFIS model with 19 rules resulting from classification	67
Figure 3.11: Confusion matrix of first iterations of the test data for SC	68
Figure 3.12: Confusion matrix of first iterations of the test data for FCM.....	68
Figure 3.13: The test results of different IMF combinations using ResNet-50 model	72

List of Tables

Table 2.1: Intensity histogram and GLCM-based features	30
Table 2.2: Characteristics of seven CNNs utilized in this study	50
Table 3.1: The accuracy values of highest five ML models for HHT-based features...64	
Table 3.2: Classification results of the three different approaches for ML	65
Table 3.3: Applied parameters for SC and FCM methods.....	67
Table 3.4: The results of evaluation metrics for each repetition of clustering methods	69
Table 3.5: The results of R , RMSE, MAE, and R^2 values for training and test data .	70
Table 3.6: Classification results of seven CNN architectures using first three IMFs of 250-ms sEMG segments	73
Table 4.1: Comparison Table of ML Studies.....	76
Table 4.2: Comparison table of the current FL studies in literature	78
Table 4.3: Comparison table with recent DL studies	84

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
HMI	Human-Machine Interaction
HCI	Human-Computer Interaction
HRI	Human-Robot Interaction
EEG	Electroencephalography
NIR	Near-infrared
EMG	Electromyography
sEMG	Surface Electromyography
iEMG	Intramuscular Electromyography
MU	Motor Unit
MUAP	Motor Unit Action Potential
TD	Time Domain
FD	Frequency Domain
TF	Time-Frequency
TFD	Time-Frequency Domain
FrFT	Fractional Fourier Transform Domain
RMS	Root Mean Square
MAV	Mean Absolute Value
PSD	Power Spectral Density

TFA	Time-Frequency Analysis
TFR	Time-Frequency Representation
WT	Wavelet Transform
GT	Gabor Transform
STFT	Short-Time Fourier Transform
HHT	Hilbert-Huang Transform
EMD	Empirical Mode Decomposition
HHS	Hilbert-Huang Spectrum
DL	Deep Learning
AI	Artificial Intelligence
ML	Machine Learning
FL	Fuzzy Logic
MF	Membership Function
CNN	Convolutional Neural Networks
LSTM	Long-Short Term Memory
SVM	Support Vector Machine
kNN	k-Nearest Neighbor
NB	Naïve Bayes
ANN	Artificial Neural Networks
GPH	Globally Parsed Histogram
EVM	Extreme Value Machine
CRNN	Convolutional Recurrent Neural Network
XAI	Explainable Artificial Intelligence
CWT	Continuous Wavelet Transform
TL	Transfer Learning
SAWT	Scale Average Wavelet Transform
XWT	Cross-Wavelet Transform

ViTHGR	Vision Transformer-Based Hand Gesture Recognition
HD-sEMG	High Density Surface Electromyography
ANFIS	Adaptive Neuro-Fuzzy Inference System
FCM	Fuzzy C-Mean
PCA	Principal Component Analysis
IMF	Intrinsic Mode Function
SC	Subtractive Clustering
IF	Instantaneous Frequency
HT	Hilbert Transform
IA	Instantaneous Amplitude
PDF	Probability Distribution Function
GLCM	Gray-Level Co- occurrence Matrix
NN	Neural Network
DT	Decision Tree
LDA	Linear Discriminant Analysis
FIS	Fuzzy Inference System
FC	Fully Connected
ReLU	Rectified Linear Unit
SD	Standard Deviation
BP	Back Propagation
CM	Confusion Matrix
ACC	Accuracy
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
FP	False Positive
TP	True Positive
FN	False Negative
TN	True Negative

RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
CV	Cross-Validation
ANOVA	Analysis of Variance
SKCV	Stratified k-fold Cross Validation
WL	Window Length
BT	Binary Tree
TKEO	Teager–Kaiser Energy Operator
R-CNN	Region-based Convolutional Neural Network
TTS	Train-Test Split
WPT	Wavelet Packet Transform
TÜBİTAK	The Scientific and Technological Research Council of Türkiye

List of Symbols

$b(t)$	Inference EMG Signal
$m_g(t)$	Sum of MUAP Trains
$\delta(t - t_{gh})$	Motoneuron Discharge Instants
$s_g(t)$	Waveform of the Action Potential of Each Single Unit
R	Number of Motor Units
T_i	Total of discharges for the g -th Motor Unit
g	Motor Unit
h	Discharge
t_{gh}	Discharge Instants
$*$	Convolution Operation
$x(t)$	Input Signal of EMD Algorithm
$e_{min}(t)$	Lower Envelope
$e_{max}(t)$	Upper Envelope
$a(t)$	Average of the Lower and Upper Envelopes
$d(t)$	Detail
$r(t)$	Residue
u	Number of IMF
H	Distribution of the Data
p	Threshold Value of the Statistical Significance of the Data

$h(t)$	Input Signal of HT
$\frac{1}{\pi t}$	Cauchy Kernel
$\hat{h}(t)$	HT of Input Signal $h(t)$
$H\{.\}$	HT Operator
π	Pi
τ	Formal Variable for Integration
$z(t)$	Analytical Signal
$A(t)$	Instantaneous Amplitude
$\theta(t)$	Instantaneous Phase
$\omega(t)$	Instantaneous Frequency
$H(\omega, t)$	3D Spectrum
G	GLCM
(x, y)	Spatial Position of Image
$I(x, y)$	Pixel Value
I	HHS image
Δx	Spatial Offset
Δy	Spatial Offset
$M \times N$	Dimension of image I
$\rho(i, j)$	Pixel Value at the Location (i, j) of an $M \times N$ Size Image
(i, j)	Pixel Location
p	Number of Offsets
m	Mean of GLCM
σ	Standard Deviation
σ^2	Variance

S	Skewness
Kr	Kurtosis
En	Energy
E	Entropy
Co	Contrast
Ho	Homogeneity
m	Important Principal Components a d -dimensional Dataset
d	Dimension of Dataset
f	Classification Discriminant
o_i	Data Value in SVM
v_j	Support Vector
s	Number of Support Vectors
Y_j	Class of Data o_i
a_j	Lagrange Multiplier
K	Kernel Function in SVM
b	Threshold Value Set
\mathcal{P}	Probability
d	Euclidean Distance between Data Point and Nearest Neighbor
K	Number of Neighbor to Selected Data Point
j	Node number of ANFIS
x_1	FIS Inputs of Node i
x_2	FIS Inputs of Node i
A_j	Linguistic Label of FIS
B_j	Linguistic Label of FIS

f	Overall Output of ANFIS
p_j	Linear Parameter of FIS
q_j	Linear Parameter of FIS
r_j	Linear Parameter of FIS
Π	Node Label of Second Layer of ANFIS
N	Node Label of Third Layer of ANFIS
W	Weight in FIS
i	Rule Number
\overline{W}_i	Normalized Firing Strength in FIS
\mathcal{X}	Dataset in SC
n	Number of samples in Dataset \mathcal{X}
x_i	Possible Cluster Center
\mathcal{D}_i	Density of Surrounding Samples
r_a	Positive Constant in SC
r_b	Positive Constant in SC
\mathcal{D}_{c1}	Maximum Density
ϵ	Certain Fraction
x_1^*	Initial Cluster Center in SC
n	Number of Vector in FCM Clustering
c_j	Cluster's Center in FCM
x_n	Data Samples in FCM
μ	Membership Matrix
μ_{ij}	Degree of membership of object j in cluster i
j	Object

i	Cluster
m	Degree of Fuzziness
d_{ij}	Euclidean Distance between c_i and x_j
J	Objective Function
U	Fuzzy Partition Matrix
l	Layer Number of CNN
p	p^{th} Row Under Consideration
q	q^{th} Column Under Consideration
k_l	Kernel Number of l^{th} Layer
$f_l^k(p, q)$	(p, q) Element of Feature Matrix
(x, y)	x^{th} and y^{th} Coordinate Under Consideration of An Image
P	Total Number of Rows of Feature Matrix
Q	Total Number of Column of Feature Matrix
c	Channel Index
I_c	Input Image Tensor
$i_c(x, y)$	(x, y) Element of c^{th} Channel of an Image
$e_l^k(u, v)$	(u, v) Element of k^{th} Kernel of l^{th} Layer
F_l^k	Convolution Output for l^{th} Layer and k^{th} Neuron
Z_l^k	Pooled Feature-map of l^{th} layer for k^{th} Input Feature-map F_l^k
$g_p(\cdot)$	Specie of Pooling Processing
N_l^k	Batch Normalization Output
T_l^k	Transformed Output
$g_a(\cdot)$	Activation Function
m_z	Mean of Output Neuron

s_z	Standard Deviation of Neurons' Output
γ	Learning Parameter of Normalization
β	Learning Parameter of Normalization
$\mathcal{H}(P, Q)$	Cross Entropy Loss Function
w	Actual Value
\hat{w}	Predicted Value
P	Objective Distribution
Q	Proximate of Objective Distribution
ν	Number of All Classes
R	Coefficient of Correlation
R^2	Coefficient of Determination
k	Actual Value
\hat{k}	Predicted Value of k
\bar{k}	Mean Value of k
f_s	Sampling Frequency

1 Introduction

1.1 Hand Gesture Recognition

State-of-the-art devices developed to facilitate and help human lives have spread to all areas of life. These areas are in a wide range from military and industrial to entertainment, and especially health. They have led to the need for effective interaction between people and themselves. Because they need to be able to perceive humans correctly to perform their functions. In order to do that, machines, computers, and robots must be able to receive correct input data from humans. This can be performed via human-machine interaction (HMI), human-computer interaction (HCI), and human-robot interaction (HRI) approaches [1]. Although they have systemic differences, they were basically developed for the same purpose. The common issue for all of them is how to ensure the interaction between humans and the system in the most effective and simple way.

Amputee prostheses, rehabilitation and surgical robots, teleoperation in surgery, haptic systems, virtual reality applications, and exoskeletons can be given as examples of the systems that are most used in the mentioned interactions. The common characteristics of these examples is that they help human movements and actions [2]. It is the most logical way to use the most effective organ in movement and communication to ensure communication between people and these systems. Because the right interaction can be achieved with the right communication tool. Hands are one of the most effective communication tools in the human body, and they are also a significant element for control in the movement-based devices [3]. In the controlling of prostheses, medical robots, virtual reality, or smart system applications as seen in Figure 1.1, providing input data from the hand to the system is the most effective and easy approach for performing the planned action or movement [4,5].

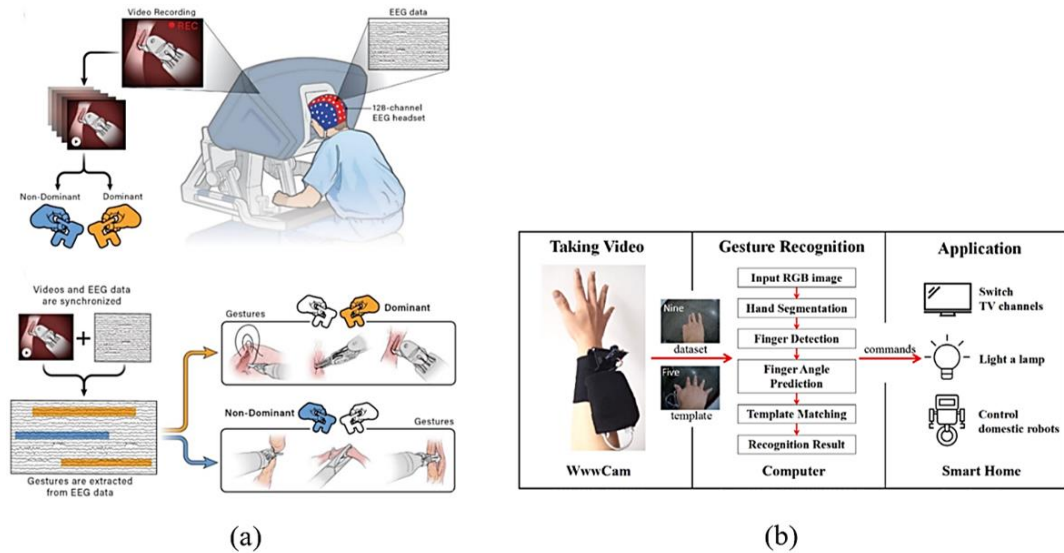


Figure 1.1: Examples of gesture-based systems: (a) EEG-based hand gesture recognition system [4], (b) vision-based hand gesture recognition system [5]

When hand gestures are used in the system interface, the user can use the device with the best performance with the help of the recognition of the desired movement by the system. Prediction of hand movement also ensures system security by minimizing latency for critical applications such as prosthetics and surgical robots [6]. Video-based, acceleration sensor-based, acoustic-based, brain-signal-based (electroencephalography (EEG)), near-infrared (NIR)-based, or muscle signal-based (electromyography (EMG)) are used to predict hand movements as seen in Figure 1.1 and Figure 1.2 [7]. In here, the use of biological signals (as seen in Figure 1.2) containing movement information is mostly used [8].

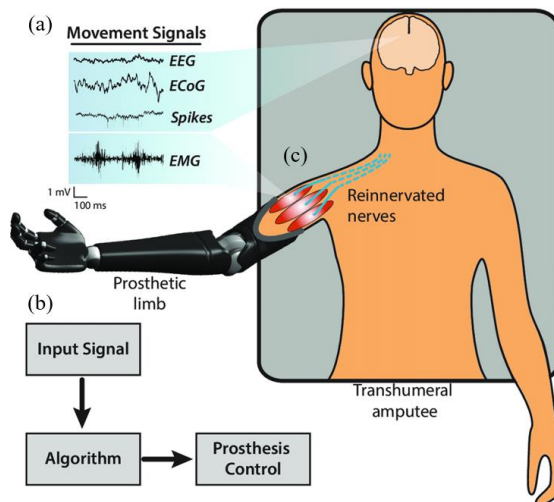


Figure 1.2: A biological signal-based hand movement recognition system: (a) prosthesis control using different biological signals, (b) controlling algorithm for pattern recognition, (c) data collection from healthy muscles of amputee [8]

Muscle signals from biological signals are the most preferred input data. Muscles from biological elements in the body are the most basic element of the movement system and are directly responsible for the realization of movement [9]. They are one step ahead of other approaches in providing information about the movement. They contain the electrical muscle activity they produce while performing the movement and the information about the movement. Therefore, because EMG data contains information about muscles, it can be an effective input data for many hand gesture recognition applications [10].

1.2 sEMG Signal and Analysis

EMG is significant and effective data for other movement-based applications like disease diagnosis besides control and rehabilitation strategies as seen in Figure 1.3 [11]. EMG signals can be collected using different approaches. While there are so many different types of data, EMG data is also divided into two groups according to the collection approach as noninvasive and invasive. The surface EMG (sEMG) is taken with electrodes placed on the skin surface as noninvasive way [12], and intramuscular EMG (iEMG) is taken with an electrode placed directly into the muscle with a needle or fine-wire as invasive way [13]. sEMG is more popular because it is a painless, non-invasive, and easily applicable method [12].

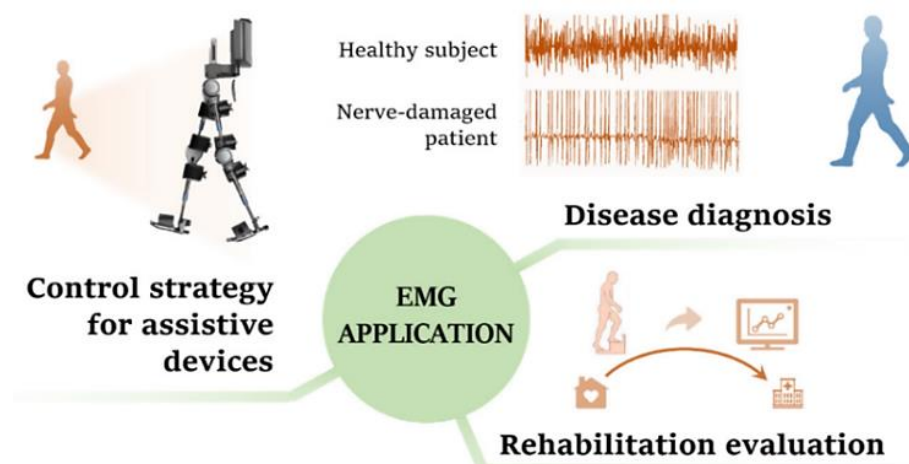


Figure 1.3: EMG applications areas [11]

1.2.1 sEMG Generation

The sEMG signal is the electrical activity generated by a contracting muscle, where there is a flow of ions. The rate of this flow is the electrical current and is measured in Amperes. The current within the muscle changes the electrical potential, and the difference in it between two locations of the muscle is evaluated in Volts. This voltage from the surface electrode is affected by the impedance or noises, and the changing its dispersion on the skin as a result of muscular activity is measured as the sEMG signal [14].

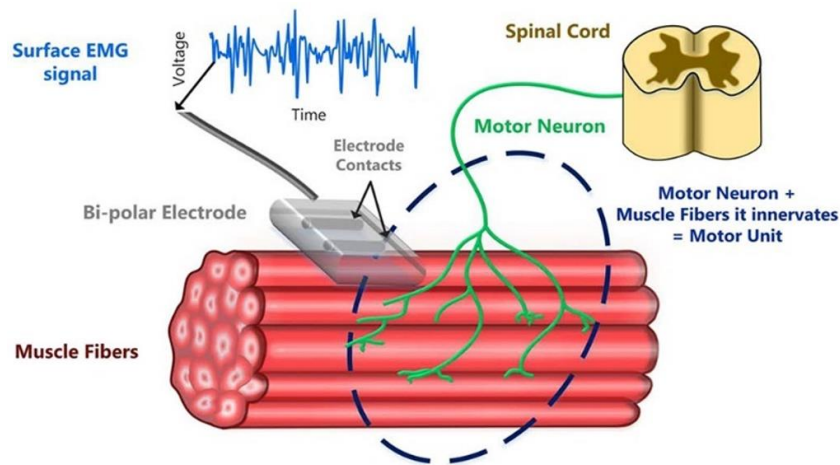


Figure 1.4: sEMG generation process [14]

Motoneurons conduct the signals from the spinal cord to the muscles. A motoneuron is started, and it innervates an action potential across the muscle fiber membrane in each fiber. The motor unit (MU) action potential (MUAP) means to the electrical potential owing to activating the fibers [6]. sEMG is a total of action potentials produced by motor units within the electrode area on the skin as seen in Figure 1.4 [14]. It is known that sEMG is used in many studies in the literature. In this thesis, sEMG data were also used.

1.2.2 Mathematical Representation of sEMG Signal

The fundamental model for the sEMG signal is defined as a sum of the MUAPs during the static contraction. EMG shows non-stationary characteristics like other biomedical signals. EMG is often defined as an interference signal due to the combine many different MU contributions. Depending on the muscle contracted and the intensity of

contraction, the number of stimulated motor units can reach huge numbers. The interference EMG signal here can be defined as in Equation (1.1):

$$b(t) = \sum_g^R m_g(t) = \sum_g^R \sum_h^{T_g} \delta(t - t_{gh}) * s_g(t) \quad (1.1)$$

where $b(t)$ is interference EMG, and consists of summation of MUAP trains, $m_g(t)$, each described as the time convolution among the discharge instants $\delta(t - t_{gh})$ and the waveform $s_g(t)$ of the action potential of each single unit. R corresponds to the number of MUs collected, and T_g is the total of discharges (h) for the g -th MU. In accordance with Equation (1.1), two principal welds clarify $m_g(t)$: the discharge instants t_{gh} and the waveform representing the MUAP, $s_g(t)$. The mathematical explanation of collecting sEMG signal is shown in Figure 1.5, where the true signal and noise compose the recorded sEMG signal [15].

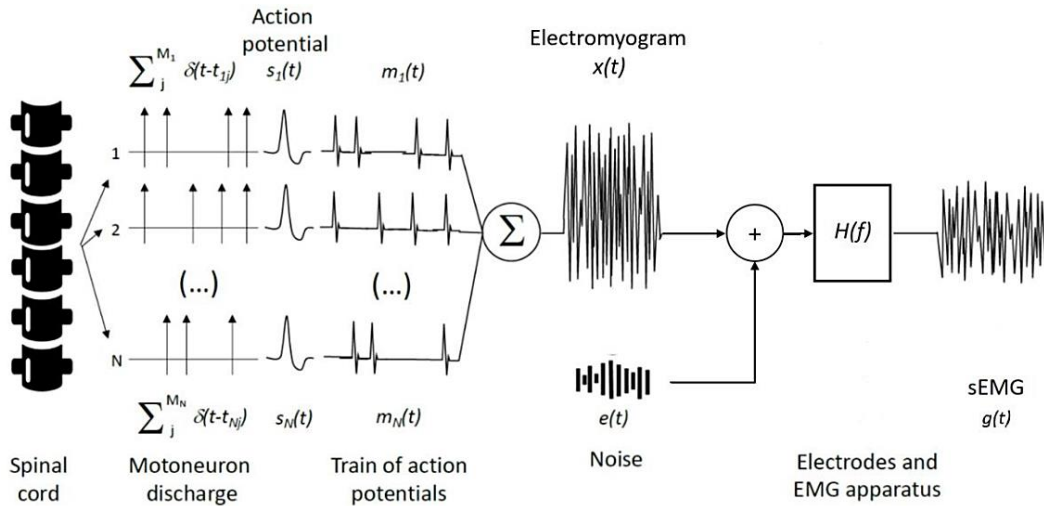


Figure 1.5: Structure of the recorded sEMG: $g(t)$ is the recorded sEMG signal on which analysis is performed, $x(t)$ is the interested true signal, $e(t)$ is noise, and $H(f)$ is the transfer function of the record tool [15]

1.2.3 sEMG Signal Preprocessing and Analysis

The biological signals need to be preprocessed and cleaned before they are used, and then various analysis approaches are applied to access the information inside them.

1.2.3.1 sEMG Preprocessing

Depending on the used approach, the processed or raw sEMG signal itself can be a direct input to the system, or it can be represented by features or images extracted from the signal. Like any biological signal, sEMG signals include noises from environment. So in order to reach the relevant information in it, the signals should be cleaned of unnecessary information and noise [16]. By this, it is more advantageous in many ways to use the noise-free, that is processed version of the signal instead of the raw form. The data that has been cleaned and brought to a certain frequency range or the new data to be obtained from this data contains real movement information. Thus, the clearest and most accurate information about the movement can be reached. Noisy EMG signals have 50 Hz interference. This needs to be cleared with the notch filter. In addition, the frequency range of the EMG signal is generally accepted between 20–500 Hz [17]. According to the application, the signals should be brought to the desired frequency range with a band-pass filter. An example procedure for this is demonstrated in Figure 1.6 [14].

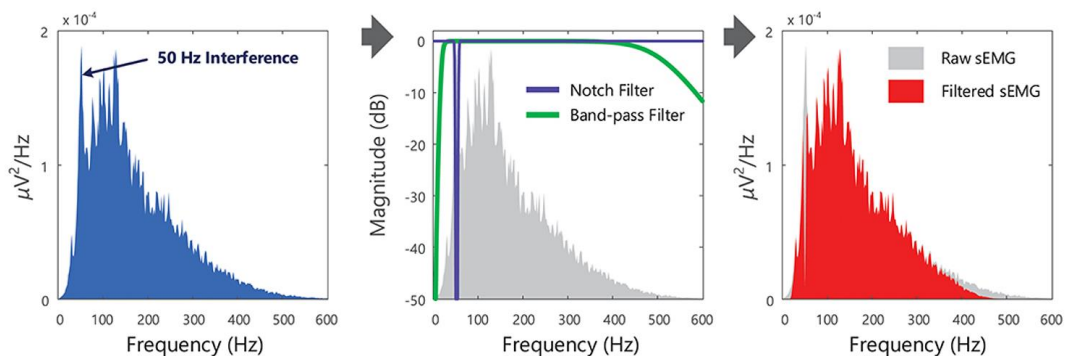


Figure 1.6: An example of a raw sEMG with 50 Hz interference, the frequency response of a 50 Hz notch filter and a 20–500 Hz band-pass filter, and the resulting filtered signal, respectively [14]

1.2.3.2 sEMG Signal Analysis

In addition to the preprocessing processes of biological signals, more in-depth analysis can be provided with analysis approaches and various information about the signal can be obtained. EMG signal analysis can be performed in different domains. The analysis or useful information form can be obtained from the time domain (TD), frequency domain (FD), time-frequency (TF) domain (TFD), or fractional Fourier transform (FrFT) domain, etc. While TD demonstrates the changing of a signal over time, FD

demonstrated the signal's energy distribution over a range of frequencies. In TD, the computational complexity is lower. Root mean square value (RMS), zero crossing, and mean absolute value (MAV) are mostly used as TD features. FD features are mostly used for muscle fatigue analysis. In FD, the power spectrum density (PSD) is mostly used [18]. The mean frequency and median frequency are examples of FD features. TFD possesses the ability to analyze varying frequency information at distinct time moments. Various time-frequency analysis (TFA) methods are applied to the signal to bring it to TFD [9].

- **Time-Frequency Analysis Methods**

The TFAs are studied in an extensive manner for distinct applications of biomedical signal processing. They make it possible to match a signal in the TD with a two-dimensional (2D) function of time and frequency. They come from a time-varying spectrum and are called TF representations (TFR). In this way, the information on the frequency spectrum in a specific time range or where the signal in a specific frequency range is placed in time can be examined. The density or energy distribution may be observed visually. Recent studies used TF images in DL architectures and have obtained successful results. The Wavelet Transform (WT) [19], Gabor Transform (GT) [20], and Short Time Fourier Transform (STFT) [3] can be given as an example of conventional TFAs. Especially, it is known that Wavelet Transform is studied in a widespread manner for different purposes in biological signal processing [19]. In Figure 1.7, an example of nonstationary signal's TD and FD forms, and TFR obtained from the WT is shown [21].

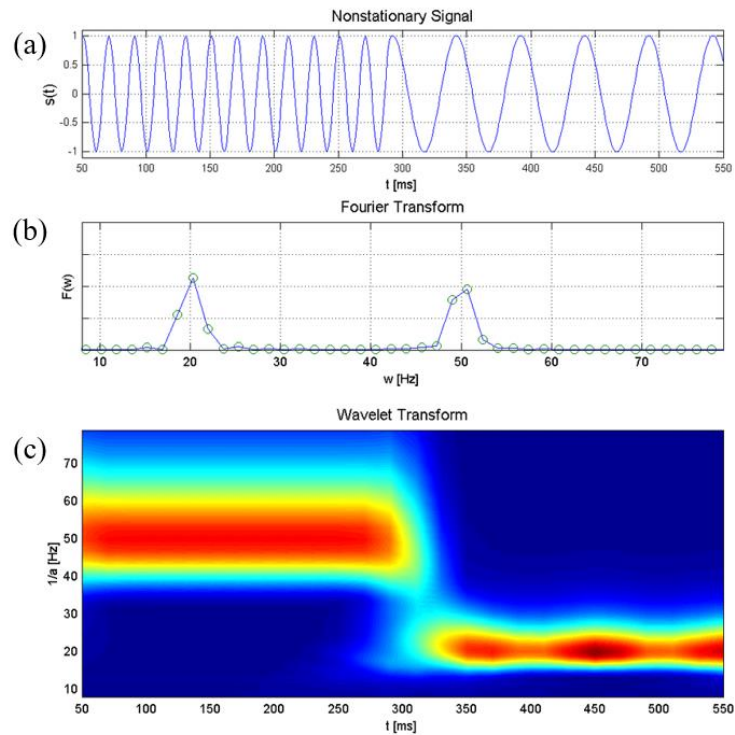


Figure 1.7: The analysis of non-stationary signal using WT: (a) in TD, (b) in FD, and (c) a time-frequency map in TFD [21]

Another TFA method, the Hilbert-Huang Transform (HHT) has been proposed to analyze non-stationary signals [22]. HHT consists of applying the Empirical Mode Decomposition (EMD) to the given signal, followed by obtaining Hilbert-Huang Spectra (HHS) of the extracted modes. The HHT is a data-adaptive approach that has important advantages in extracting the most appropriate features of short-time EMG signals for further processing. HHS images, an advanced representation of EMG in the joint TF plane, benefit from the advantage of utilizing 2D images in DL. Similarly, in this study, we evaluate the advantages of HHT in providing a high-resolution TF image over traditional methods such as the WT and STFT with its data-adaptive structure and multi-resolution ability [23]. In addition, the HHT-deep learning (DL) combination performs better than the STFT-based method for mining non-stationary features [24].

1.3 Artificial Intelligence Approaches

In order to provide interaction between human and devices, the processed EMG data must be classified in artificial intelligence (AI) models. There are several AI approaches for EMG classification. Machine learning (ML), deep learning, and fuzzy logic (FL) are the one of the AI approaches. These three approaches can be categorized in AI as seen in Figure 1.8.

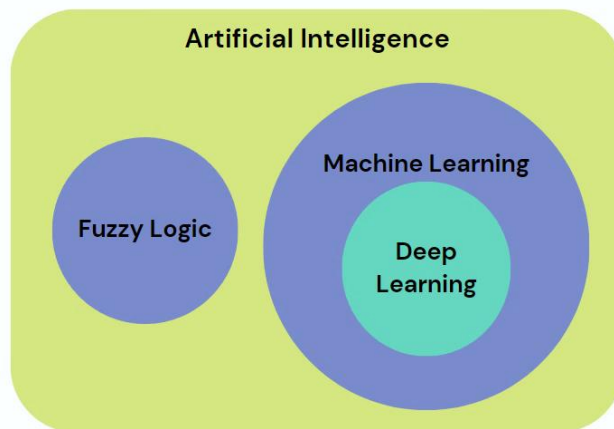


Figure 1.8: The schema of artificial intelligence methods based on this study

Machine learning is one of the basic approaches in AI methods for EMG-based gesture classification. EMG is non-stationary, and ML is able to produce a remedy for its non-stationary operations [25]. In order to be able to classify with ML models, feature extraction performed by an expert is required. A feature can be defined as a function of one or more measurements that specifies some measurable property of an object or image. The features obtained from the signal are also effective data representations. Features represent the classes they belong to and can be used as inputs to classifiers. ML models that use extracted features as input, although require an extra step and expert intervention, are more reasonable than other AI methods due to their computational cost and speed [6]. For this reason, it is often preferred because it provides classification success and low complexity advantages at the same time.

Similarly, fuzzy logic for the classification of EMG of hand movements operates a similar process as ML. FL algorithms also need some numerical values, namely features, that accurately represent the EMG signals. The fuzzy logic system can classify a model of biological signals such as EMG. Unlike other approaches, it may be the preferred approach for EMG-based movement classification, with its operation

most similar to the human thought system [26]. For example, if the feature of a myoelectric signal is explained with a daily expression such as 'slightly high', the said value can be expressed and classified with the fuzzy membership values in the fuzzy set model. However, the fixed membership functions (MF) in fuzzy logic delimit information more in the rule base than MF base. So, FL can suppose more memory and processing time [27]. In addition, there are more advanced systems consisting of a combination of fuzzy logic with neural networks. The features in several domains can be extracted from one-dimensional (1D) signals or from 2D images obtained by the conversion of EMGs, and are used in ML and FL.

In addition to FL and ML, deep learning approaches, which have been very popular in recent years for many tasks, are also used in EMG classification. In addition to classification, DL is also used for many applications such as medical image segmentation and object detection. Besides, Convolutional Neural Networks (CNN) demonstrate outstanding performance in biomedical tasks with large image-based datasets [28]. DL can automatically extract features in their deep networks from data representations, eliminating the need for extra feature extraction and selection steps. For this, the generally preferred method is to express EMGs as images and classify them in deep network models. In addition, deep networks can use both the signal itself (like in Long-Short Term Memory (LSTM)), its pre-extracted features [29], and images as input data. Two-dimensional rather than one-dimensional representations of signals can be obtained by transforming EMGs into images. It is known that 2D representations improve classification performance thanks to the processing of spatial information [28,30]. Considering all this, DL can provide more advanced and highly accurate results than other methods. Therefore, it is very important to obtain the correct image representation of EMG signals and use it in deep networks.

1.4 Hand Gesture Classification Studies in the Literature

The studies in the literature are listed under three headings within the scope of the topics of this thesis: machine learning-based, deep learning-based, and fuzzy logic-based studies. An example of hand gesture recognition system architecture is shown in Figure 1.9.

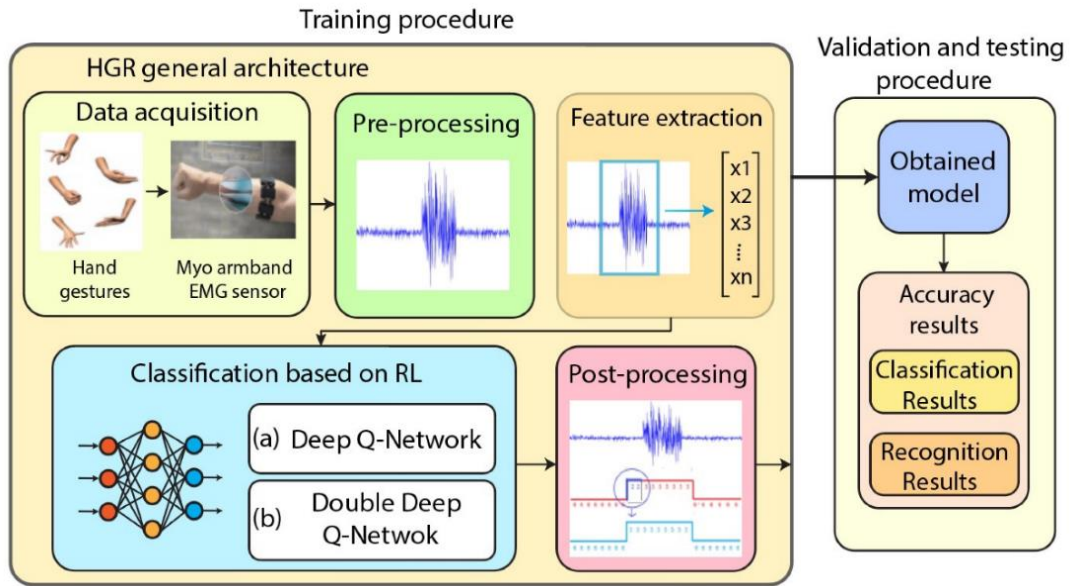


Figure 1.9: An example schema of EMG-based hand gesture recognition system [31]

1.4.1 Machine Learning-based Studies

ML-based classification approaches are mostly used in hand gesture classification. Tepe and Demir [32] performed real-time gesture classification of seven gestures. Several features were extracted to use in the support vector machine (SVM) model. They reached an accuracy of 95.83% for personalized data and reached the best average classification accuracy of 91.79% with mean absolute value for the generalized dataset. Briouza et al. [33] performed the comparison of two ML models, which were k-nearest neighbors (kNN) and SVM, using different feature combinations. From five TD features, they reached the best classification accuracy of 78.56% in the SVM model using MAV, waveform length, and slope sign change. Nunez-Montaya et al. [34] proposed a classification approach to anthropomorphic robotic hands. They extracted 20 features from EMG signals and classified them into eight ML models. The highest accuracy of 84.78% was obtained by the multi-layer perceptron model. Karapinar Senturk and Bakay [35] classified seven gestures from UCI 2019 EMG dataset. Eight-channel data were classified in six different ML models, and Naïve Bayes (NB) obtained the highest accuracy of 96.43%. Gouda et al. [36] proposed a two-stage classification architecture to classify seven gestures. The architecture was composed of SVM and artificial neural networks (ANN) stages, and hand-crafted features of 2-channel EMG from the time domain and time-frequency domain were classified in it. They reached an average accuracy of 99.00%. Alguner

and Ergezer [37] aimed to prove entropy could be used only one feature for real-time classification using different sliding window lengths. They used Ninapro DB5 and SingleMyo datasets in different ML models. They reached the best validation accuracy with SVM. Also, the proposed globally parsed histogram (GPH) approach increased accuracy from 60.35% to 89.06%. Rabin et al. [38] classified six movements using the STFT features. They used kNN after implementing the PCA and obtained an accuracy of 77.30%. Devaraj and Nair [39] used the features from TD of seven movements. They reached an accuracy of 93.00% with the kNN. Shi et al. [40] classified TD features in the kNN, and reached a classification accuracy of 94.00%. Azhiri et al. [41] used Extreme Value Machine (EVM) to classify the reflection coefficients and obtained a classification accuracy of 91.00%.

1.4.2 Deep Learning-based Studies

DL-based studies can vary according to the input data. Toro-Ossaba et al. [42] used pre-processed EMG signals as a direct input to the LSTM-RNN model. They performed only preprocessing and cleaning of EMGs. They used 4-channel EMG of 5 gestures and reached an accuracy of 87.70% in real-time tests. Kim et al. [43] performed real-time hand gesture recognition to provide HRI. They reached an accuracy of 96.00% using the convolutional recurrent neural network (CRNN) for 10 dynamic gestures in real-time. Gopal et al. [44] presented a comparison study, which compared ML and DL models. 12-channel EMG of selected 10 gestures from the Ninapro dataset was used. They performed feature extraction from the time domain for ML-based models and used sEMG images for the DL model. They reached that DL and ensemble models were better than ML models. But for all cases, their classification accuracy remained below 90.00%. Gozzi et al. [45] performed the explainable AI (XAI)-based classification study. They aimed to obtain a deeper understanding of the importance of the hand-crafted ML features and features from the DL model. While 91.72% accuracy was obtained with SVM, they obtained values up to 97.00% with CNN. They proved that even with some complexity issues, the DL model is preferable to ML. Karnam et al. [46] proposed a hybrid CNN and Bi-LSTM-based EMGHandNet for hand activity classification. They used five different databases and obtained the highest accuracy of 93.48% for six gestures with the UCI Gesture dataset. Nahid et al. [47] used two different EMG datasets to classify gestures

and control the prosthesis. Firstly, they applied Continuous Wavelet Transform (CWT) to EMGs to create images and then classified them in transfer learning (TL) and CNN-LSTM-based models, like ResNet, AlexNet, etc. ResNet-18+LSTM model reached the highest accuracy of 99.83% for ten finger movements with the Rami Khusaba dataset. Gunes and Akkaya [48] classified 2-channel EMG signals of six hand movements. They applied CWT to the signals to convert them into scalogram images. Also, they performed different channel strategies during classification and classified channel-1 data and channel-2 data in GoogLeNet, separately. They reached a higher accuracy of 97.22% with channel-1. Oh, and Jo [49] classified 8-channel sEMG of six movements. They applied STFT, WT, and Scale Average Wavelet Transform (SAWT) to convert signals into images to give CNN. SAWT images achieved the best classification accuracy of 93.90%. Huang and Chen [50] performed a combination of CNN and LSTM to classify STFT spectrograms of EMGs. They reached a classification accuracy of 80.93%. Roy et al. [51] utilized the Cross-Wavelet Transform (XWT) spectrums in the CNN. They reached an accuracy of 97.60% for four gestures. Montazerin et al. [52] proposed a recognition architecture, which is vision transformer-based hand gesture recognition (ViTHGR). In order to test their model, they used the 128-channel High Density-sEMG (HD-sEMG) data of 65 gestures. They fed their model with three-dimensional (3D) images and their model reached the test accuracy of $84.62 \pm 3.07\%$.

1.4.3 Fuzzy Logic-based Studies

FL-based studies are not as common in the literature as ML and DL studies. However, various approaches have been tried for fuzzy-based gesture classification over the last decade. While the studies of the last 3 years for ML and DL were examined, the studies in the last decades for fuzzy are summarized in this section. Khezri and Jahed [53] designed a multi-stage sEMG pattern recognition system using a functional technique with improved performance at each step for prosthesis control. The hand movement commands were defined by the Adaptive Neuro-Fuzzy Inference System (ANFIS) using sEMG signals. Time and time-frequency domains and their combinations are used as features of the sEMG. The proposed recognition model with ANFIS classification using combination features achieved an average classification accuracy of 92.00%. Balbinot and Favieiro [54] investigated the development and study of a

system that uses sEMG to characterize specific movements of the upper extremity. In order to recognize seven hand and arm movements, RMS values in the feature dataset obtained from 8-channel EMG of 30 subjects were used as input in a pattern recognition algorithm based on neuro-fuzzy, and the average accuracy value was obtained as 86.00%. Jahani Fariman et al. [26] proposed a technique that provides a simple and computationally efficient classification. Features of sEMG signals from 2 muscles of 4 participants were extracted using a new combined time-domain feature extraction method. The fuzzy C-mean (FCM) clustering method and scatter plots were used to evaluate the performance of the proposed multi-attribute method against other proven multi-features. As a result of the classification of features in ANFIS and ANN, ANFIS showed higher classification accuracy of 88.90% with significant improvement in computation time. Control with sEMG signals can also be used in systems such as wheelchairs. Based on the need for low-cost wheelchairs for people with physical disabilities, Kaiser et al. [55] designed a solar-powered wheelchair. Raw sEMG signals of the arm muscles were used to move the wheelchair and the features extracted from the raw sEMG signals were classified by ANFIS and achieved 96.85% accuracy. Ulkir et al. [56] performed the extraction of the RMS, waveform length, and kurtosis features from the two-channel EMG and classified them using a Mamdani-type fuzzy logic-based classification method. The classification process for two hand gestures was performed with an average accuracy of 92.13%. Caesarendra et al. [57] used ANFIS to classify five finger gestures. It reduced 16 features to three using principal component analysis (PCA) and achieved an average classification accuracy of 72.00%. Arozi et al. [58] performed the extraction of features from the signals to provide input for the prosthesis, and the features were classified in ANFIS. They reached an accuracy of 98.09%. With such studies, it has been emphasized that sEMG and fuzzy logic classification approaches can be used to perform wrist flexion/extension, forearm supination/pronation, and hand opening/closing movements of prosthetic devices.

1.5 Aim of the Study

The aim of this study is to perform hand gesture classification by applying the TFA approach to three different artificial intelligence methods. When classifying hand gestures in ML and FL methods, it has been seen that the features of TD or FD are usually extracted from the signal and used in classification studies. There are also studies using TFD or other domain features. In addition, in recent years, it has been determined that images from TFA have been used in DL models. From this point of view, it is aimed to classify the EMGs of hand movements using the same data type (TFA-based) in all three artificial intelligence approaches and to evaluate the results. Since the ML and FL models cannot receive the image directly, an approach is planned in which the features extracted from the TFR images are input to these models, and the TFR images are given to the DL model as input. Hilbert Huang Transform was chosen as the TFA approach. TFA-image-based features are used for the first time in ML and FL, and HHT images are used for the first time in the DL models.

2 Materials and Methods

In this study, three distinct AI approaches were applied to classify EMG data of hand gestures. In order to make an evaluation of AI methods, the TFA-based data is used in all three AI models. Hilbert-Huang Transform is used as a TFA technique. Depending on the structure of ML and FL models, hand-crafted features are extracted from TFA images and are given to models, and the HHT image data is directly given to the DL model. The schema of this thesis is shown in Figure 2.1.

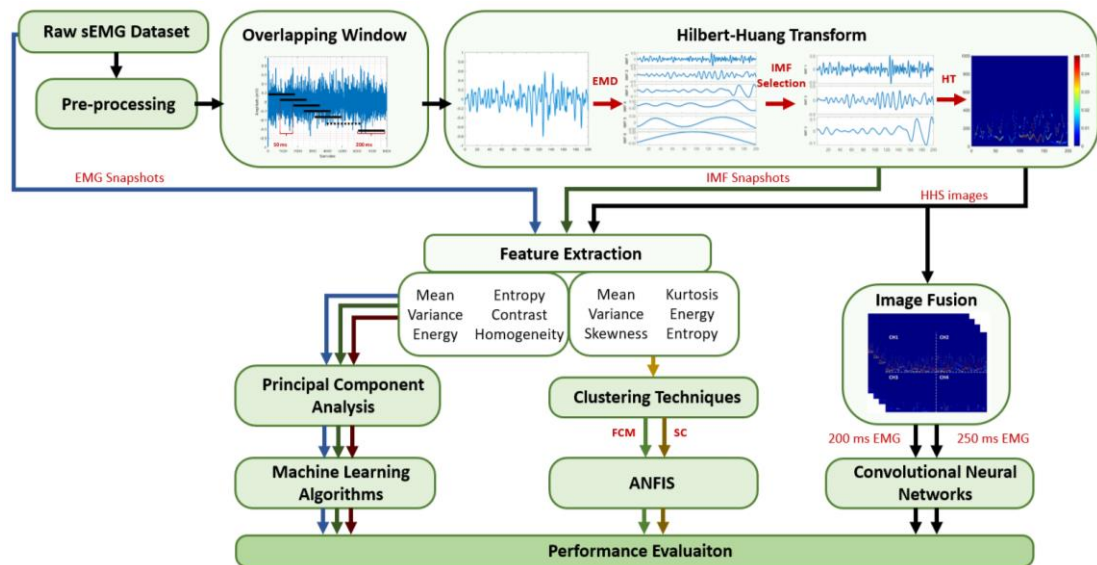


Figure 2.1: The schema of the method

In here, HHT-based features and images are used as common data to evaluate the performance of the three main AI models. AI models have been evaluated within themselves. In order to evaluate machine learning in itself, classification is made using three different feature types. These are the TFR image-based features separately extracted from EMG snapshots, the intrinsic mode functions (IMF) snapshots, and the Hilbert-Huang spectrum images. The image features extracted from HHS images are classified using two different clustering techniques, which are Fuzzy C-mean and

Subtractive clustering (SC), for the fuzzy logic model. In the deep learning-based method, the images are compared by classifying them in different CNN models using different lengths of EMG segments, which are 200 ms and 250 ms. In addition, the performance of different IMF combinations is evaluated.

2.1 sEMG Data

Since this thesis is a continuation of the undergraduate graduation project, the data collected and published as a dataset during the undergraduate thesis process were used. The mentioned publicly available sEMG dataset [16] is used in this thesis. The data was collected considering the Helsinki Declaration, and ethical permission was granted by İzmir Kâtip Çelebi University. In the dataset (Version 1), there are signals from 30 healthy subjects, 15 female and 15 male considering equal gender distribution. An average age of 22.37 ± 1.47 was calculated for all subjects. The average weights of the female and male are 57.03 ± 5.39 kg and 69.6 ± 9.3 kg, and the average heights are 164 ± 6.13 cm and 173.67 ± 5.96 cm, respectively. sEMG signals were collected from the right hands in any case of the dominant hand, and three participants are left-handed. All participants performed seven hand movements; the rest of the hand, the extension of the wrist, flexion of the wrist, ulnar wrist deviation, radial wrist deviation, punch (not too tight or loose), and fingers abduction (open hand), respectively as shown in Figure 2. 2.

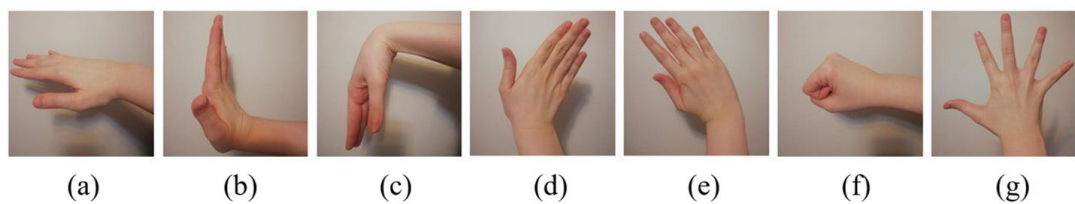


Figure 2. 2: Seven hand gestures: (a) Rest position, (b) Extension of wrist, (c) Flexion of wrist, (d) Ulnar deviation of wrist, (e) Radial deviation of wrist, (f) punch (grip), (g) open hand (abduction)

The MP36 model BIOPAC[®] device was used for the data recording. the signals were recorded at a sampling frequency of 2 kHz using four Ag/AgCl surface bipolar electrodes. The approximate location of these channels, where each represents a muscle, is shown in Figure 2.3-(a). The muscles are flexor carpi radialis, extensor carpi radialis, flexor carpi ulnaris, and extensor carpi ulnaris as seen in Figure 2.3-(b).

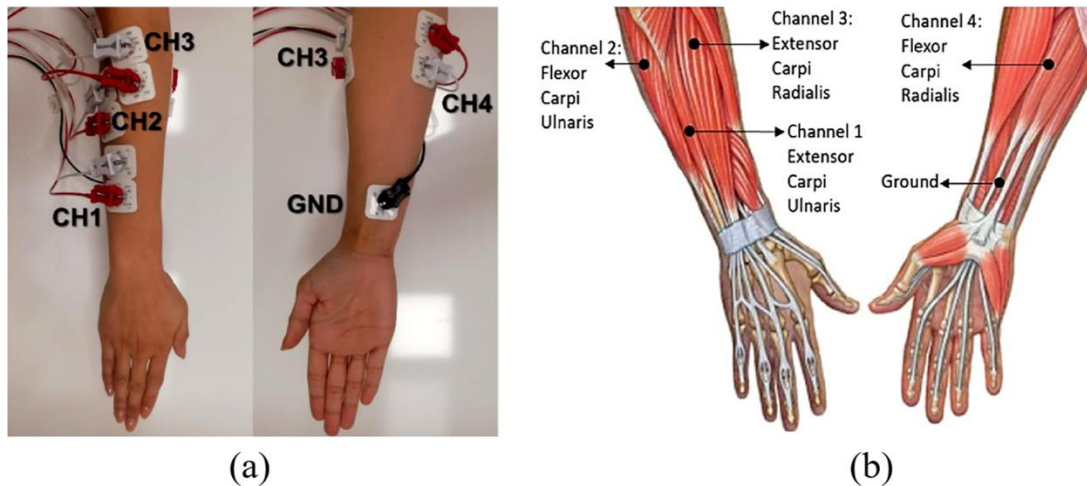


Figure 2.3: Location of electrodes for posterior and anterior views of right arm: (a) 4-channel and ground electrodes placement, and (b) muscles corresponding to electrode positions

During the recording, the subjects performed seven hand movements with 6 s duration, and 4 s resting times in between gestures. In Figure 2.4, the timeline is demonstrated, where a cycle length is 74 s. A cycle begins with a 4 s rest. When the 6 s movement period started, each gesture was performed once and stayed in a steady state till the end of the movement period. Before taking the next gesture, there is a rest period of 4 s again. Each cycle was repeated 5 times, with 30 s resting periods between them. The whole length of the signal is 490 s for each subject.

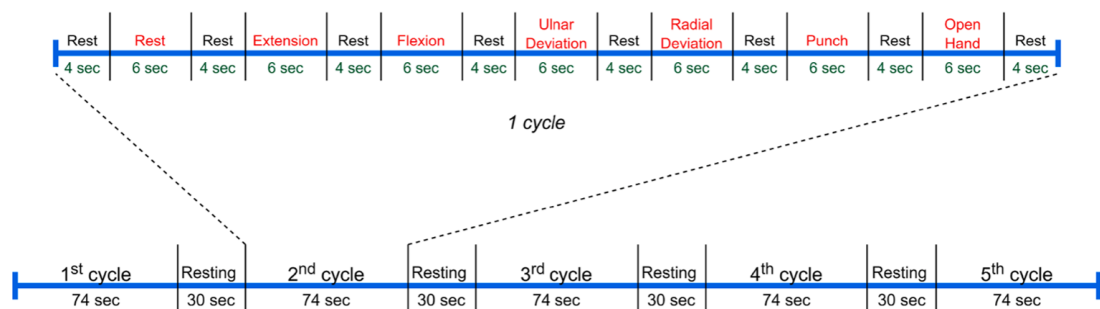


Figure 2.4: Timeline of data collection procedure

2.2 Pre-processing

Analysis of EMG signals of movements consists of processing the signals, feature selection, determination of the classification algorithm and the obtained prediction accuracy. The first stage of this is the filtering and segmentation processes performed in the preprocessing stage.

2.2.1 Filtering

While sEMG is one of the most effective methods of detecting hand movement, it suffers from some essential electrical noise from the environment such as electrical devices. It is also influenced by other biological signals originating from neighboring muscles or internal organs [59]. To extract precise information about a movement from the sEMG signal, filtering processes are applied to the signals, thereby removing irrelevant information [60]. Accordingly, a sixth-order Butterworth band-pass filter of 5–500 Hz is applied to the recorded signals to eliminate the artifacts, and a second-order notch filter at 50 Hz is applied to minimize the power-line interference [9]. Signals cleared in this way are taken for use in FL and DL [61,62]. In addition, in ML, a different approach is applied, and based on the view that the dominant EMG frequency was mostly above 50 Hz [63,64], a sixth-order Butterworth band-pass filter of 50–500 Hz, and a second-order notch filter at 50 Hz are applied to the same raw sEMG signals [65].

2.2.2 Segmentation

Segmentation is used to describe periods of time when the muscle contracts and rests, which is performed to extract relevant and informational regions from the signal. In addition, it is very important in terms of replicating the data in the dataset and evaluating the method for use in real-time applications. Segmentation is performed by taking the 4 s steady-state period in the middle by not taking the 1 s transient states at the beginning and the end of the 6 s signal as seen in Figure 2.5. Because, it is considered that these periods, where the muscles stayed in the greatest contraction status as the essential sEMG of the concerned gestures. In similar studies [66], the steady-state periods provide more successful results because most of the myoelectrical activity occurs in those times. Thanks to this, the possible delays (owing to the transition) can be eliminated that can happen at the end and at the beginning.

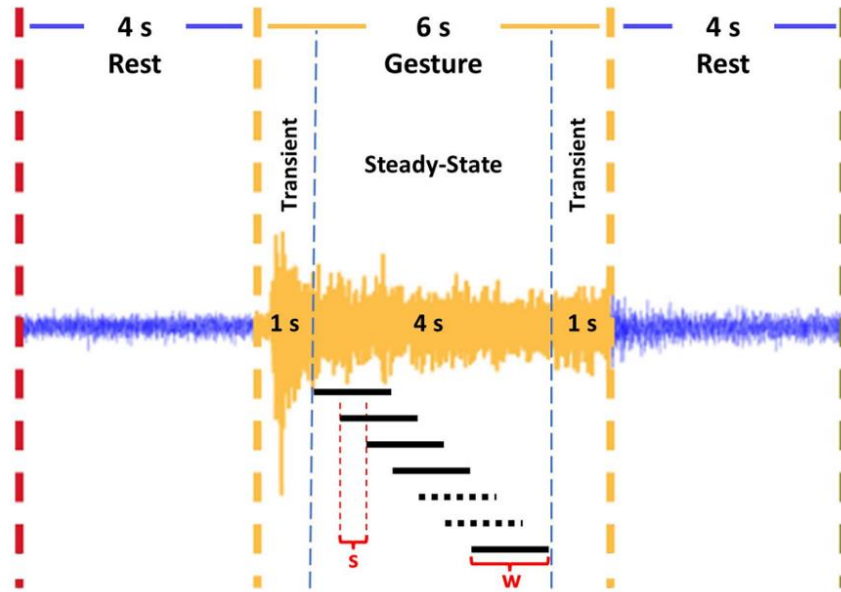


Figure 2.5: Overlapping window process for steady-state EMG signal (s: shift length, w: window length)

Relatively smaller window sizes (≤ 250 ms) are chosen to demonstrate robustness and feasibility in online and real-time applications, such as to achieve lower variance in feature extraction steps and provide effective control of hand prostheses [67]. During the windowing phase, 200 ms windows with 75% overlap (150 ms overlapping) were used for ML and FL applications. For DL, two different windows, 200 ms and 250 ms, with 75% overlap (150 ms overlapping), were used to determine the optimum parameters and to examine the TF image-based classification performance. All operations were done in the MATLAB[®] R2021a (The MathWorks Inc., USA) environment. These processes are visualized in Figure 2.5, where s represents the shift length (50 ms) and w represents the window length (200 ms and 250 ms). A 200 ms segment obtained from a long 4 s steady state signal is shown in Figure 2.6.

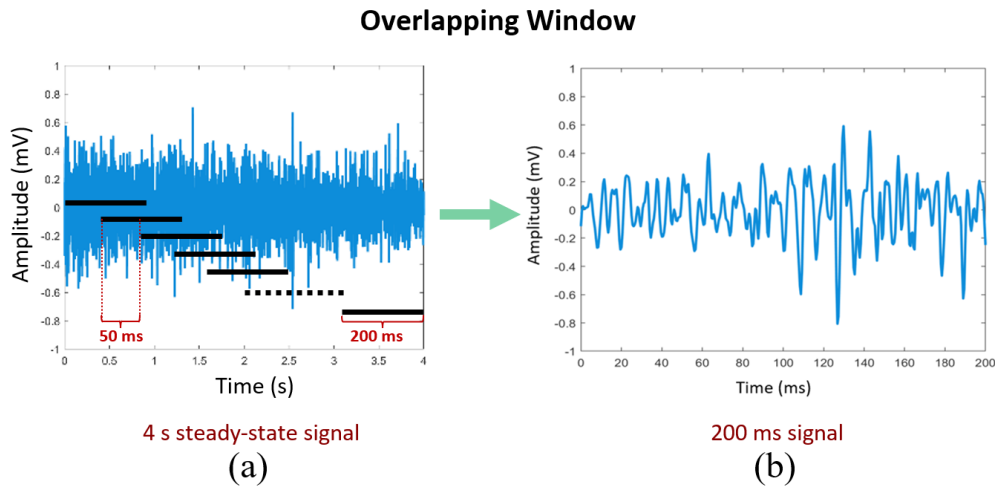


Figure 2.6: Overlapping window process: (a) 4 s steady-state EMG signal, and (b) 200 ms EMG segment

2.3 Time-Frequency Analysis

TFA techniques ensure to the investigation of the biological signal and represent it in the TF domain. TFA methods are used to acquire the distribution of signal energy over the TF domain. This provides the transformation of 1D biosignals into 2D TF images, which are a more efficient representation [51]. TFA allows us to determine when various signal frequencies are present by calculating a spectrum, usually at regular time intervals. When the signals are analyzed at certain time intervals, the frequency spectrum of the signals can be observed better. It is known that the spectral properties of the signal can be better studied in time and frequency planes if the signal analysis is performed using time and frequency-limited functions. TFA methods, which are frequently used in signal processing, allow matching a signal in the time domain with a function of time and frequency. These methods result in time-varying spectrum representations [22].

TF images can be convenient input for creating DL models or TF image features that contain both time and frequency information can be used in ML-based models. The state-of-the-art studies that used TF images in the CNN architectures have given successful results [68,69]. Various transformation techniques such as STFT, Fourier Transform (FT), WT, and Gabor transform are used in the signal analysis of distinct areas [20]. Considering the stationary and non-linear nature of biological signals, the correct information about the characteristic of the signal can be obtained by the

instantaneous frequency (IF) value, and this value can be best determined by Hilbert Transform (HT) and the characteristic information of the signal can be detected [22].

In this study, the HHT method is used as a TFA technique. The performances of ML, FL, and DL models are investigated. To use the new data obtained as a result of TFA in all AI models, TF images are then used in the pre-trained CNN models, and TF image-based features are used in ML and FL models. It is briefly introduced the HHT method in the following.

2.3.1 Hilbert Huang Transform

HHT is an adaptable TFA technique utilized to acquire time-varying frequency characteristics of multi-component non-stationary biological signals. It supplies more meaningful and extremely localized signal representation in the TF domain. This representation is specifically called TFR, which is a result of the TFA techniques. Firstly, the application of the EMD algorithm to the signal allows the signal to be decomposed into intrinsic mode functions and instantaneous frequency information to be learned. Each IMF is a non-predefined function. IMFs are extracted from the EMG by the agency of the “sifting” algorithm in a data-driven way. The HT is applied to the selected IMFs to provide the IF and instantaneous amplitude (IA) of the modes. Then they are utilized for the composition of 3D TF spectrums as Hilbert Huang Spectrums [22]. In addition, the selection of IMF, which is an intermediate process in order to determine which of the IMFs resulting from EMD will apply HT, can also be considered as a part of this transformation.

2.3.1.1 Empirical Mode Decomposition

EMD ensures to decompose of the non-linear and nonstationary signal into a limited number of modes named IMFs [2,70]. Each IMF, which are primary oscillation function, should satisfy the following situations:

- The zero crossings and extrema numbers should be equal or should differ at most by 1.
- The upper and lower envelopes’ means should be 0.

The sifting process of EMD of the input signal $x(t)$ is explained below steps:

Step 1. Acquire local maxima and minima values of the signal

Step 2. Interpolating the local extrema places, calculate both lower $e_{min}(t)$ and upper $e_{max}(t)$ envelopes

Step 3. Calculate $a(t)$ that is the average of the lower and upper envelopes

$$a(t) = [e_{min}(t) + e_{max}(t)]/2 \quad (2.1)$$

Step 4. Compute the detail, $d(t)$, by subtracting the average $a(t)$ from the input signal $x(t)$

$$d(t) = x(t) - a(t) \quad (2.2)$$

Step 5. The detail signal is checked to see if it fulfills the requirements of being an IMF or not. If so, $d(t) = IMF_1(t)$. If not, return to **Step 1**, and perform again the following steps substituting $d(t)$ for the input signal.

Step 6. After calculating $IMF_1(t)$, compute the residue $r(t) = x(t) - IMF_1(t)$. If the calculated $r(t)$ owns more than one zero-crossing, return to the first step and calculate another IMFs till the stop criteria are accomplished [71]. All algorithm results in an addition of IMFs plus an $r(t)$ term that can be utilized to rebuild the signal as follows:

$$x(t) = \sum_u IMF_u(t) + r(t) \quad (2.3)$$

IMFs here are referred to as first-order IMFs. The whole unbundling process ends with a limited number of IMFs. The highest frequency oscillation in the EMG signal is symbolized by the first IMF, and the following contains lower frequency oscillations of the signal. The final $r(t)$ shows only the general trends of the signal [72]. The first six IMFs are shown as an example in Figure 2.7.

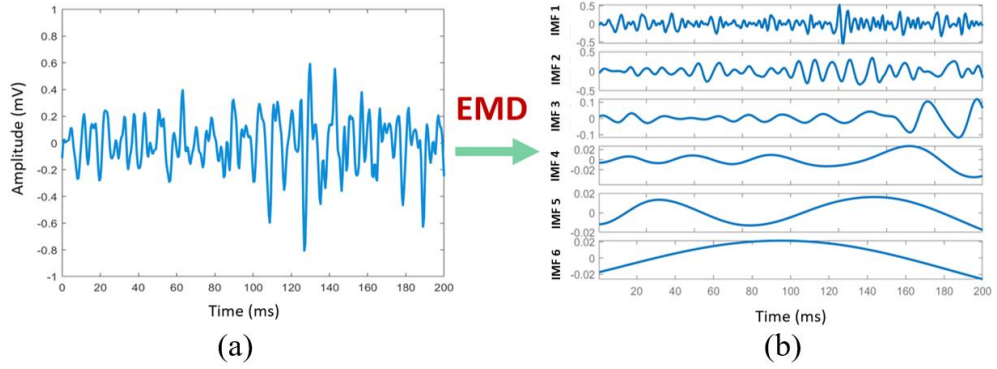


Figure 2.7: Implementation of EMD to the EMG signal, (a) 200 ms EMG signal, and (b) the visualization of obtained first six IMFs

2.3.1.2 IMF Selection

The selection of suitable IMFs for Hilbert Transform contributes positively to movement classification success [73]. There are many methods for IMF selection of a biological signal. Methods based on correlation, energy, probability distribution function (PDF), power spectral density, and statistical significance using t-tests are frequently used for the selection [74]. Among these methods, the method based on statistical significance provides effective results and is a suitable approach for EMG. In this method, in order to examine the statistical significance, the H value indicates whether the distribution of the data is normal or not, and the p -value (<0.05), which is accepted as the threshold specifying the statistical significance of the data. These values are calculated. When the p -value of the IMF is calculated, if it is greater than the threshold, the IMF has statistical significance ($H = 0$), otherwise, the IMF does not have a normal distribution ($H = 1$) [73]. In this study, the p and H values of each IMF are calculated and ordered from largest to smallest. According to this order, three IMFs meeting the significance condition are selected. Figure 2.8 shows an example that the first three IMFs that meet the significance condition. HHS images that are created with different combinations of the most significant three IMFs are classified in DL architectures and the best performance is determined.

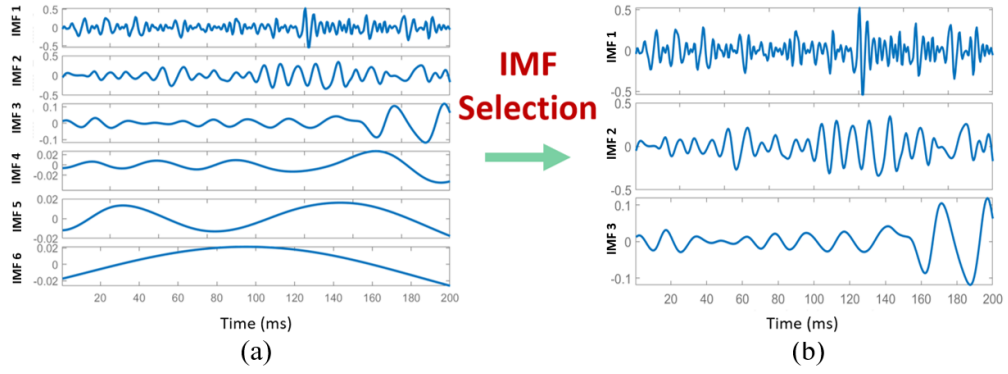


Figure 2.8: Statistically significant IMFs selection: (a) the first six extracted IMFs, and (b) the selected first three IMFs

2.3.1.3 Hilbert Transform

Hilbert Transform is the method used to obtain an analytical equivalent of a given signal and to demodulate signals. It is possible to create a 3D graph (time-frequency-energy) containing time and frequency information. It is defined as the convolution of a given signal $h(t)$ with $\frac{1}{\pi t}$, given by:

$$\hat{h}(t) = H\{h(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{h(\tau)}{t - \tau} d\tau \quad (2.4)$$

where $H\{\cdot\}$ operator indicates the HT. τ is the formal variable for integration process. The analytical signal $z(t)$ corresponding to $h(t)$ is obtained by:

$$z(t) = h(t) + j\hat{h}(t) = A(t)e^{j\theta(t)} \quad (2.5)$$

where $A(t)$ and $\theta(t)$ correspond to the amplitude and the phase functions, respectively. The analytical signals possess one-sided spectra and can be utilized to describe the signal's instantaneous properties as follows:

$$A(t) = [h(t)^2 + \hat{h}(t)^2]^{1/2} \quad (2.6)$$

$$\theta(t) = \arctan(\hat{h}(t)/h(t)) \quad (2.7)$$

$$\omega(t) = \frac{d\theta(t)}{dt} \quad (2.8)$$

In the mentioned definitions, $A(t)$, $\theta(t)$, and $\omega(t)$ are instantaneous amplitude, instantaneous phase (IP), and instantaneous frequency, respectively [33]. After performing all steps for each IMF, a 3D spectrum $H(\omega, t)$ is created placing $A(t)$ at the appropriate location $(\omega(t), t)$ in the TF plane [75]. HT operation for selected IMFs is shown in Figure 2.9. In summary, the HHS image, $H(\omega, t)$, is obtained by the following steps:

- Estimation of the instantaneous properties of each selected IMF ($A(t)$; $\theta(t)$; $\omega(t)$).
- Generating the 3D representation of the amplitude placed in the TF plane, $H(\omega, t)$.

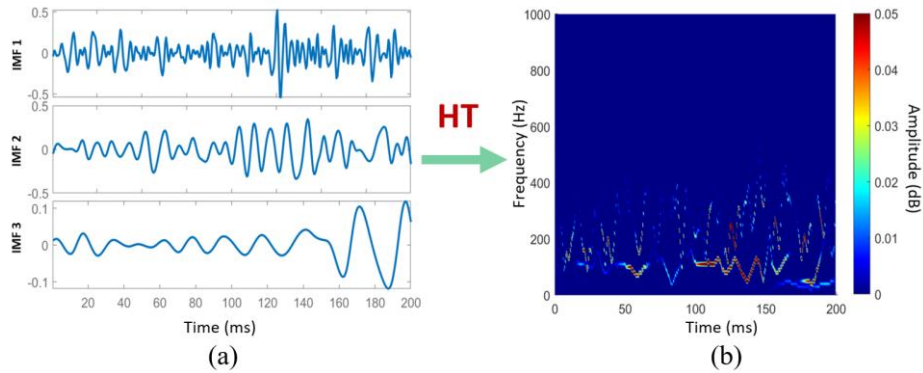


Figure 2.9: HHS generation: (a) the selected three IMFs, and (b) Hilbert-Huang Spectrum after application of Hilbert Transform

2.3.2 TF Image Fusion and Snapshot Image Generation

TFA processes for deep learning-based models are applied to 200 ms and 250 ms segments, separately. HHS images are obtained from each signal segment for four EMG channels. For the HHT implementation, the generated 3D HHS can be represented as a volume or a 2D image utilizing color mapping. That means were projected into 2D images. Hence, the instantaneous energy of segments turns into a color representation in the 2D image. The color scale of entire images is normalized before the fusion to compose an equitable classification. The four channels

representing muscle were combined with the image fusion approach before being given to the DL-based models.

In order to compare the performance of the features extracted from the HHS images for ML models, a snapshot of the 200 ms sEMG segment obtained as a result of windowing and a snapshot of three IMFs selected in the IMF selection stage are taken to extract image features from them.

2.4 Feature Extraction for ML and FL

A feature can be defined as a function of one or more measurements that specify some measurable pattern of an object or image. FL and ML methods require an additional feature extraction step from data before classification. This additional step aims to symbolize the signal or data in an informative and also minimal form. The classification of TFR images in an AI model calls for improved hardware. To accomplish the issue, distinct techniques are recommended to symbolize TFRs in a simplistic and low dimension condition.

Efficient features represent specific characteristics of the classes they belong to and can be used as inputs to classifiers. In this thesis, gesture classification is performed utilizing the features extracted from the visualized HHS images. If the image features that can be obtained by various computational methods are carefully selected, they can present the characterization of the image and represent the maximum relevant information required for its analysis. In this study, features based on intensity histogram [76] and gray-level co-occurrence matrices (GLCM) [30] are used to symbolize HHS images. The GLCM approach [77], which calculates the occurrence of gray-level intensities in neighbor pixels, is utilized as a statistical approach to extract texture features from medical images. In detail, it performs measurement of the frequency with which pixels in a set direction and margin are existing in the image. Also, it investigates their spatial intercourse. GLCM composes a square matrix. The dimension of this matrix is equal to the gray levels' number in the related image. All cells of GLCM correspond to the count of the co-occurring concerned gray levels [30]. The calculation of the GLCM matrix is given in Equation (2.9):

$$G_{\Delta x, \Delta y}(i, j) = \sum_{x=1}^M \sum_{y=1}^N \begin{cases} 1, & I(x, y) = i, \text{ and} \\ & I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (2.9)$$

where I is a HHS image with the dimension $M \times N$, and (i, j) is the pixel value. (x, y) is the spatial position in the image I . $I(x, y)$ is pixel value. Δx and Δy are spatial offsets, which takes 0 and 1 in this study, respectively. In GLCM, each pixel's information is acquired with distance and angle orientations. Too close can produce homogeneity, while too far can produce information among pixels to be irrelevant. Therefore, calculating according to the angles determined by keeping the distance value at the optimum level provides reliable results [78].

An example GLCM process is simply shown in Figure 2.10. In Figure 2.10, the distance between the interested pixel and that pixel's neighbor is defined as a p -by-2 matrix of integers, where p is the number of offsets. All rows in the matrix are two-element vectors, [row offset, column offset], where they are rows number and columns number between interested pixel and neighbor, respectively. The vector defines the relation, of a pair of pixels. The offset is frequently stated as an angle. According to the example where distance is taken as default from 'graycomatrix' command in MATLAB[®], the GLCM calculation process is visualized for four angles, which are 0°, 45°, 90°, and 135°.

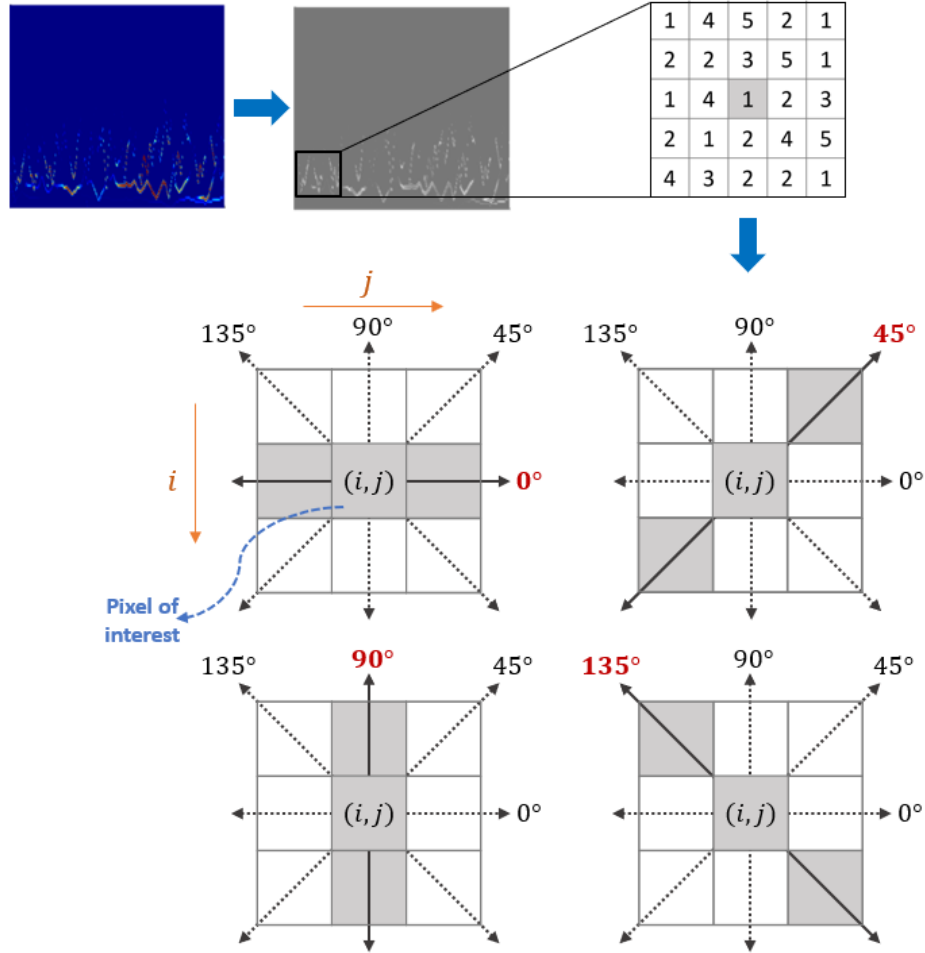


Figure 2.10: The visualization of GLCM process

Extraction of features using GLCM and texture-based approaches ensures advantages in the classification of biomedical data [30,79]. When the image analysis approaches in the previous studies are investigated, it has been seen that it includes various statistical features such as mean, variance, entropy, etc. [80] In this study, intensity histogram and GLCM-based features are selected to use. For FL, mean, variance, skewness, kurtosis, energy, and entropy are calculated to classify. In ML, mean, variance, energy, entropy, contrast, and homogeneity features are used. ML and FL used four features in common, but two features were chosen differently. Assuming the pixel value $\rho(i, j)$ at the location (i, j) of an $M \times N$ size HHS image, the mentioned features can be calculated with the formulas shown in Table 2.1. The mentioned ML features are calculated for HHS images from 200 ms EMGs, snapshots of 200 ms EMGs, and selected IMFs extracted channel-wise EMD. The mentioned FL features are also calculated from HHS images from 200 ms EMGs.

Table 2.1: Intensity histogram and GLCM-based features

Feature	Definition	Formula
Mean (m)	It defines the average intensity level of an image as the sum of the pixel values divided by the number of pixels.	$m = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \rho(i, j)$
Variance (σ^2)	It is the square of the standard deviation (σ), representing the variation of intensities around the mean.	$\sigma^2 = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (\rho(i, j) - m)^2$
Skewness (S)	It allows to characterize the degree of asymmetry of the pixel distribution around the mean value of the histogram.	$S = \frac{1}{MN} \frac{\sum (\rho(i, j) - m)^3}{\sigma^3}$
Kurtosis (Kr)	It measures the peak value or flatness of the data relative to the normal distribution.	$Kr = \frac{1}{MN} \frac{\sum (\rho(i, j) - m)^4}{\sigma^4}$
Energy (En)	Energy is defined as uniformity and is in the range [0 1]. Returns the sum of the squares of the GLCM properties.	$En = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \rho^2(i, j)$
Entropy (E)	It is defined as a measure of how irregular or random an image is.	$E = - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \rho(i, j) \log_2 \rho(i, j)$
Contrast (Co)	It specifies the measurement of the drastic alteration in gray level among neighboring pixels.	$Co = - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \rho(i, j) (i - j)^2$
Homogeneity (Ho)	It represents the likeness in gray level among neighboring pixels.	$Ho = - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \frac{\rho(i, j)}{1 + (i - j)^2}$

2.5 Classification in AI models

The classification step is aimed to use the HHS images and features in three AI approaches. In addition, each method is compared within itself by applying different approaches. The strategies followed in the method for the classification phase are explained below and the publications obtained separately by using each method during this study with their references are given:

- For ML: Six image features extracted from HHS images of 200 ms EMG signals are classified in ML models. The classification performance of image-based features extracted from EMG snapshots and IMF snapshots in ML is also evaluated [65].
- For FL: Six image features extracted from HHS images of 200 ms EMG signals are classified into FL-based architectures by applying two different clustering techniques, SC and FCM [61].
- For DL: Different combinations of IMFs for HHS image generation are evaluated. HHS images of 200 ms and 250 ms EMG signals are classified separately in seven different popular deep network architectures [62].

2.5.1 Machine Learning

ML is defined as predictive-based modeling. Varied ML methods are utilized to classify seven hand gestures. Before whole ML algorithms in the Classification Learner App (CLA) are tested with features, the feature selection method is applied to reduce dimensions and achieve the best performance. The principal component analysis method is applied to six features before the classification to represent whole features with the lowest variance.

2.5.1.1 Feature Selection

Data should be normalized before feature selection and dimensionality reduction. Data at different scales can cause misleading components. Normalization rescales the data between 0 and 1. In this work, z-score normalization is applied to the feature matrix so that each feature is included with a convenient effect.

When working with data, problems such as redundancy, overfitting and multicollinearity may occur during training. The model needs to work with optimum time and performance. To overcome the mentioned problems, feature selection and dimensionality reduction methods can be applied. In variable selection, the variable in the dataset is preserved or completely removed, while the number of variables is reduced by creating new variables consisting of a combination of existing variables in size reduction. PCA is a multivariate statistical analysis technique used to solve the aforementioned problems. The basic logic behind PCA is to represent multidimensional data with fewer variables by capturing the key features in the data. Hence, it provides the dimensionality reduction of the predictor space.

PCA transforms a multivariate system consisting of many interrelated variables into a system consisting of fewer and unrelated new variables as linear functions of these variables, which can also largely explain the total change of the previous system. Each of the principal components obtained in PCA versus initial variables is a linear combination of the original variables. Therefore, each basic component contains a certain amount of information from all variables. Thanks to this feature, PCA provides size reduction by using the first m important principal components instead of a d -dimensional dataset. If the first m principal components explain most of the total variance, the remaining $(d - m)$ principal component can be neglected. Compared to classical variable selection techniques, information loss will be minimized with this method [65].

2.5.1.2 ML Models

ML consists of four types of algorithms: supervised, semi-supervised, unsupervised, and reinforcement. For classification tasks, supervised learning is mostly preferred. In this type of algorithm, data contains wanted inputs and outputs, and using this data ML model develops an approach to specify how to reach inputs and outputs. ML classification algorithms in MATLAB[®] CLA, which includes 29 different techniques like SVM, kNN, neural networks (NN), NB, Decision Trees (DT), Random Forest, linear discriminant analysis (LDA), and ensemble models, are utilized in this step. Most of them are supervised learning algorithms. LDA is a simple and efficient method

that has attracted significant notes in recent years. Linear Discriminant and Quadratic Discriminant are applied as an example of that model.

SVM [81] is the most popular and effective ML approach utilized in gesture recognition. It is based on the logic of finding a hyperplane in a high-dimensional space that maximally separates the classes. It projects the low-size data to the high-size feature area using kernels. The suitable hyperplane represents the biggest division or margin between the classes. Therefore, a hyperplane chosen which has the distance to the nearest data dot (support vectors) on all classes is maximized. It is called a maximum-margin hyperplane as shown in Figure 2.11. The algorithm tends to neglect the outliers and detects the ideal hyperplane that maximizes the margin. Its maximization provides stability to the model. In the SVM algorithm, first, the data in low space are transformed into higher dimension space with the non-linear transform methods. Then, the classification of data can be performed in accordance with linear separability [82]. The classification discriminant (f) is demonstrated in Equation (2.10).

$$f(o_i) = \text{sgn} \left[\sum_{j=1}^s y_j a_j K(v_j, o_i) + \mathbf{b} \right] \quad (2.10)$$

where o_i is the data value to classify. v_j is the support vector, and s is the support vector's number. y_j is class of data v_j , which takes 1 or -1 as categories. a_j is the Lagrange multiplier. K is the kernel function that returns a number, which is the similarity metric, among the two vectors. According to that, 1 and 0 mean identical and different, respectively. The state is an inner product process, and the \mathbf{b} threshold value set is for classification. Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, and Coarse Gaussian SVM are applied as an SVM model in this study.

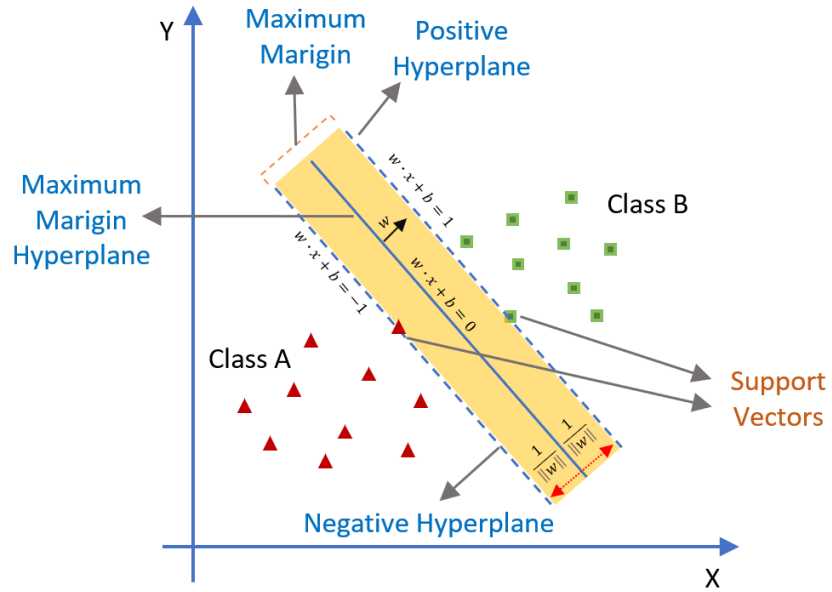


Figure 2.11: An example visualization of classification of data using SVM

The Naïve Bayes method [83] is a simple algorithm based on Bayes' Theorem that can work on unstable datasets. It is a probabilistic classifier, and it can provide robust independency suppositions among the features. In addition, the distribution of features is significant for its classification performance. Utilizing Bayesian probability, Equation (2.11) can be written for NB classifier:

$$\mathcal{P}(\mathcal{A}|\mathcal{B}) = \frac{\mathcal{P}(\mathcal{B}|\mathcal{A})\mathcal{P}(\mathcal{A})}{\mathcal{P}(\mathcal{B})} \quad (2.11)$$

where $\mathcal{P}(\mathcal{A}|\mathcal{B})$ is the posterior probability, and $\mathcal{P}(\mathcal{B}|\mathcal{A})$ is called the likelihood or conditional probability of \mathcal{B} given \mathcal{A} . $\mathcal{P}(\mathcal{A})$ is the prior probability of \mathcal{A} in which no information of \mathcal{B} takes into account. $\mathcal{P}(\mathcal{B})$ is the prior or marginal probability of \mathcal{B} that behaves as a normalizing constant [84]. Gaussian NB and Kernel NB models are selected to implement for gesture classification.

DTs are decision-backing hierarchic and tree-like models. In a DT, internal nodes symbolize the features, and each branch and each leaf node symbolize the decision rules and the outcome, respectively. One of the nodes of DTs is the decision node, which is utilized to perform all decisions and possesses various branches. Another type of node is the leaf node, which is the decisions' output and does not possess any more branches. It uses a graphical presentation to show the whole probable solutions to the related situation. Due to their structure, these types of algorithms are called decision

trees. The tree is split into subtrees depending on simple “Yes/No” questions [85]. Fine Tree, Medium Tree, and Coarse Tree are used to gesture classification.

The kNN method [86] is another popular biomedical signal classification tool. A class is assigned to the data according to which class the nearest neighbors of the data in the vector formed by the independent variables belong. The Euclidean distance (d) of each data point is calculated to define the nearest neighbor, this is shown in Figure 2.12. The aim is to obtain the least Euclidean distance. When the K value, which is the considered neighbor number, rises, accuracy goes up. For the kNN classification, a threshold value is computed by the mean of the K data dot, which is the nearest. The classification performance depends on distance, similarity, and threshold [87]. Fine kNN, Medium kNN, Coarse kNN, Cosine kNN, Cubic kNN, and Weighted kNN are applied as an example of that model.

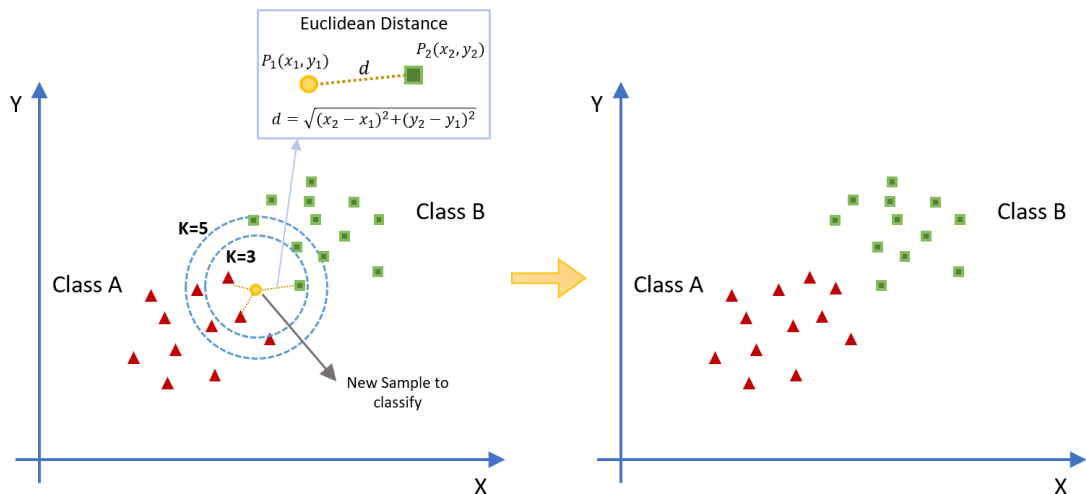


Figure 2.12: kNN algorithm: the distance between the selected point (yellow) and the neighbor points is calculated with the Euclidean distance formula, and the class is determined according to the proximity and abundance of the neighbors

A neural network is a series of algorithms that seek to identify relationships in a dataset via a process that mimics how the human brain works. A neural network is formed by many artificial neurons. These neurons are referred to as units that are composed of a series of layers. Each layer is attached to each other. In these linkages, there are specific weights to define the impact of each unit on others. The model performs the learning process as the data progresses through the units within the network [88]. There are three types of layers, which are input, hidden, and output layers. Input layers admit input data in several formats from outside. Then, the accepted data pass through one

or various hidden layers, where the transformation of input into a valued data format is performed. It calculates the weighted sum by taking the information of entire neurons in the past layer. It transmits them to the further neurons. By appointing distinct weights to each input value, the input effect is optimized and performance improvement is made. Finally, the output layer ensures the final output of the model as a result of input data. Along with these, some parameters such as the number of hidden layers and neurons, epochs, etc. affect the performance of ANN [89]. An example ANN model is shown in Figure 2.13. Narrow, Medium, Wide, Bilayered, and Trilayered NN models are used to classify.

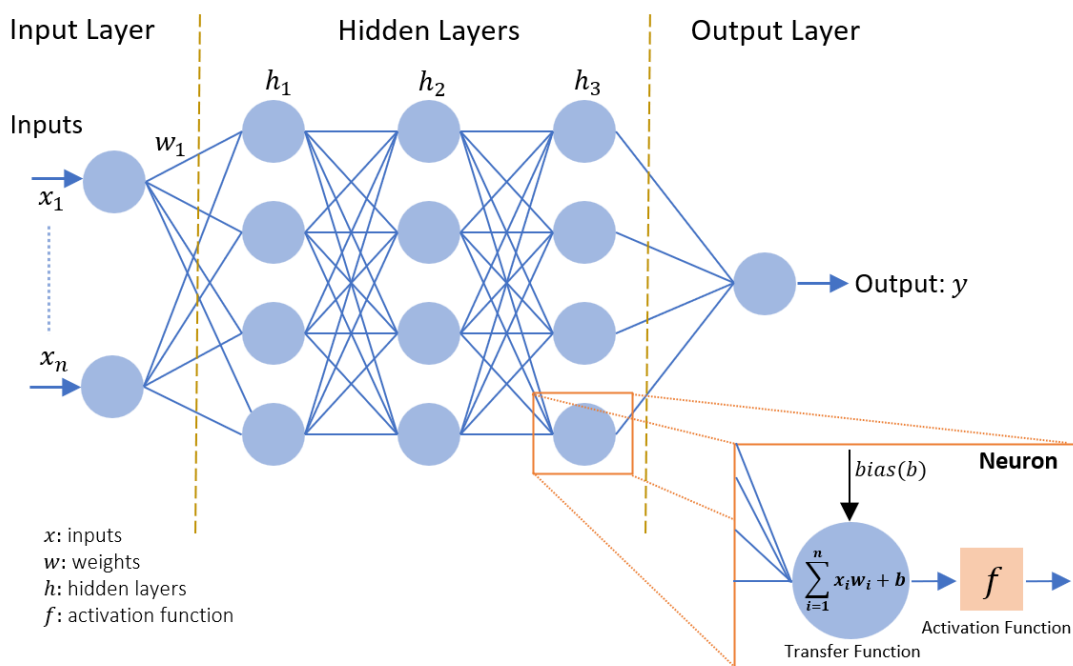


Figure 2.13: An example of ANN structure with three hidden layers

Ensemble methods utilize more than one learning algorithm to acquire preferable classification performance. It is expected to be better than the result of any algorithm using it alone. Ensemble Boosted Trees, Ensemble Bagged Trees, Ensemble Subspace Discriminant, Ensemble Subspace kNN, and Ensemble RUSBoosted Trees are used in this study.

2.5.2 Fuzzy Logic

FL may perform the extraction of unrepeatable EMG features and imitate the intention for decision-making. Especially in prosthetic hands, it has been seen that FL-based classifiers are included due to their effect on movement classification processes [90]. The fuzzy inference system (FIS) was proposed by Zadeh in 1965 [91]. FL imitates the human brain's decision system in the form of if-then rules. There are two commonly used types of FIS, Mamdani, and Sugeno. In FIS, the inputs are blurred to a value in the range [0,1]. The state of belonging or not belonging to the cluster is gradual and characterized by the fuzzy membership function used to explain it and interpreted according to the if-then rules [55].

2.5.2.1 Adaptive Neuro-Fuzzy Inference System

ANFIS is a multi-layered network structure that utilizes neural network learning algorithms and fuzzy logic. In other words, it can be said to be a feed-forward type ANN based on Takagi-Sugeno type FIS. It was first proposed in 1993 to provide the initial fuzzy system and mapping of the current input space to the output space [92]. It has an advanced structure for modeling nonlinear time series. The fact that Sugeno is more efficient and computationally efficient is one of the main reasons why Sugeno-type FIS is used instead of Mamdani in ANFIS. When the resulting value is a linear equation, it is called a first-order Sugeno FIS. If the result is a constant coefficient, it is called a zero-degree Sugeno FIS. Assuming that the system contains two inputs x_1 and x_2 , and the rule base two fuzzy if-then rules of Takagi-Sugeno type, the representation of the rules can be expressed as:

$$\begin{aligned} \text{Rule}_1: & \text{IF } (x_1 = A_1) \text{ AND } (x_2 = B_2) \\ & \text{THEN } f_1 = p_1x_1 + q_1x_2 + r_1 \end{aligned} \quad (2.12)$$

$$\begin{aligned} \text{Rule}_2: & \text{IF } (x_1 = A_2) \text{ AND } (x_2 = B_2) \\ & \text{THEN } f_2 = p_1x_1 + q_1x_2 + r_2, \end{aligned}$$

where $\{p_j, q_j, r_j\}$ are linear parameters in the conclusion part of the FIS. x_1 and x_2 are net inputs of node j ; A_j and B_j are linguistic labels. f is the overall output. The architecture of ANFIS consists of five layers shown in Figure 2.14.

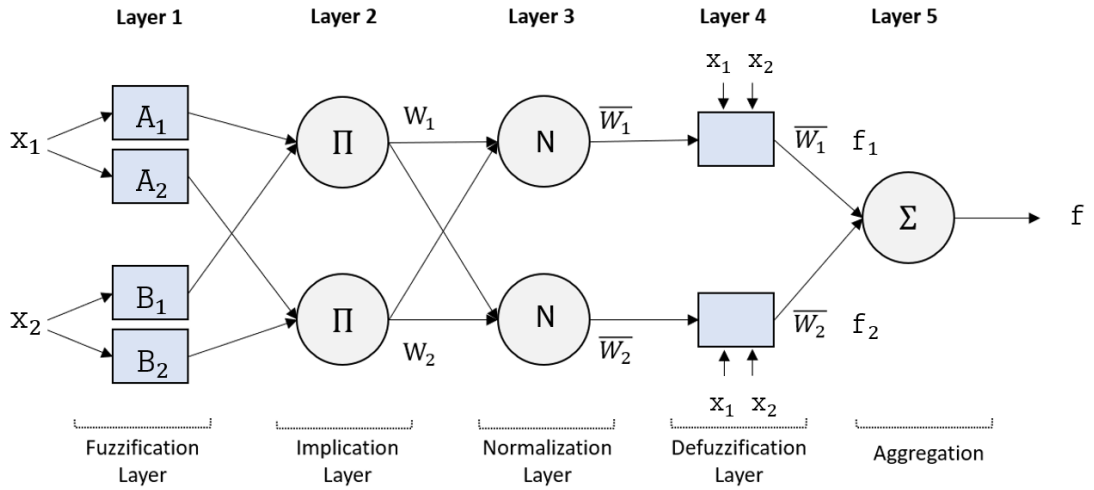


Figure 2.14: The structure of ANFIS model

Figure 2.14 shows the ANFIS structure. Each layer contains several nodes defined by the node function. Adaptive nodes represented by rectangles represent parameter sets that can be adjusted in these nodes, while fixed nodes represented by circles represent parameter sets fixed in the system. Output data from nodes in previous layers are inputs in the current layer [53]. Accordingly, what the layers mean is explained in detail below [93].

Layer 1. Establishes the degree of membership of a linguistic label. Each node of this layer is an adaptive node with the membership grade of fuzzy set.

Layer 2. The node of this layer is labeled as Π . Calculation of the firing strength of each rule, which is the representation of node output, by multiplication, is done. Here, any T-norm operator, which makes a fuzzy AND, can be utilized as the node function.

Layer 3. The node of this layer is labeled as N . Calculation of the ratio of the firing strength of rule i is performed. This layer's output is named normalized firing strength, \bar{W}_i .

$$\bar{W}_i = \frac{W_i}{W_1 + W_2}, i = 1, 2 \quad (2.13)$$

Layer 4. Calculation of the contribution of each rule i to the overall output is performed with the operation of $\overline{W}_i f_i$.

Layer 5. Lastly, calculates the overall output as the sum of each rule's contribution. The overall output is shown in Equation (2.14):

$$\text{Overall Output} = f = \sum_{i=1} \overline{W}_i f_i \quad (2.14)$$

ANFIS implements a hybrid learning algorithm, which is a combination of “gradient descent” and “least-squares” methods, to update model parameters. Each epoch of this hybrid learning process consists of a forward and backward pass. In the forward pass of the hybrid learning procedure, the loop goes forward until the output, layer 4, and the resulting parameters are defined by the least squares method. In the backward pass, the error signal propagates backward and the antecedent parameters are updated with gradients [92]. The mentioned training step is the number of iterations in the training of ANFIS. This number varies according to the dataset studied to reach fault tolerance, which is a training stop criterion related to the size of the error. In this study, ANFIS architecture is trained using Fuzzy Logic Toolbox in MATLAB[®]. Here, two clustering techniques, SC and FCM, are applied and the parameters of the ANFIS architecture are determined.

2.5.2.2 Clustering Techniques

Clustering is a data segmentation process that offers many benefits when predicting and analyzing specific problems. Different clustering algorithms are used to automatically generate Gaussian membership functions. Two of these algorithms are fuzzy C-means and subtractive clustering. Although different in implementation, both techniques create fuzzy if-then rules. SC has the learning capability that can solve complex problems without formulating them. In addition, while determining the number of clusters and cluster centers, it offers a fast and one-pass algorithm. FCM, on the other hand, provides high-accuracy results for the dataset and informs that a data point may belong to more than one cluster [90,94]. Considering the success of the SC in difficult problems and the success of FCM in data where precise boundary

determination is difficult, and considering similar studies [26,53], it promises success in gesture classification. For this study, two clustering techniques are discussed.

- **Subtractive Clustering**

SC is a fast and one-way technique for estimating the number of clusters in a dataset and the location of cluster centers. It is an extension of the mountain clustering technique and makes sense to apply this method when the number of clusters cannot be determined precisely. In the SC method, each data point is assumed to be a potential cluster center, and a measure of the probability of each point in determining the cluster center is calculated [94]. If a specified point has many neighboring points, it is considered a point with high density. In this case, the density of all points is calculated and the point with the highest density is defined as the first cluster center. This process continues by updating the density of the points so that the points adjacent to a cluster center do not become a new center. The SC function returns the cluster center in a matrix, and the rows in this matrix contain the position information of the cluster centers. This function, which also provides the number of clusters, also returns a vector containing sigma values, which defines the range of influence of a cluster center.

For mathematical explanation, suppose that dataset \mathcal{X} consists of n samples expected clustering, $\mathcal{X} = (x_1, \dots, x_n)$. It is supposed that every sample x_i is a possible cluster center and has the surrounding exemplars' density (\mathcal{D}_i) following Equation (2.15):

$$\mathcal{D}_i = \sum_{z=1}^n e^{-\alpha \|x_i - x_z\|^2}, \alpha = \frac{4}{r_a^2}, \quad (2.15)$$

where r_a is a positive constant. If x_i possesses many neighbor exemplars, it'll possess a high density. It is needed to compute the density of whole exemplars and the data dot with the maximum density \mathcal{D}_{c1} is chosen to happen the initial cluster center x_1^* . In order to refrain the neighbor data dots of the cluster center be a recent center, the whole exemplars' density should be updated following:

$$\mathcal{D}_i = \mathcal{D}_i - \mathcal{D}_{c1} e^{-\beta \|x_i - x_1^*\|^2}, \beta = \frac{4}{r_b^2}, \quad (2.16)$$

where r_b is a positive constant, and in general higher than r_a . It is repeated the above steps of acquiring a recent center and replacing densities till $\mathcal{D}_{ck} < \epsilon \mathcal{D}_{c1}$, where ϵ is a certain fraction [94]. An example application of the output of SC is shown in Figure 2.15.

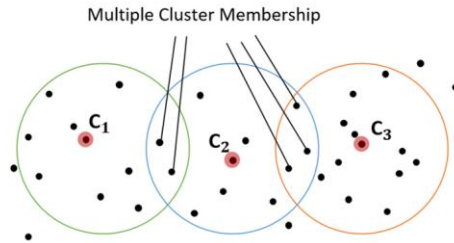


Figure 2.15: An example subtractive clustering result of data

Cluster predictions from there can be used to start ANFIS. SC parameters are set from the parameter screen in the ANFIS interface in MATLAB[®]. While applying the SC method, the *range of influence*, *squash factor*, *accept ratio*, and *reject ratio* parameters are used. The range of influence indicates the radius of a cluster, while the squash factor is used to suppress the potential of the far points of the cluster, which is used to multiply the radius values that determine the neighborhood of a cluster center. Accept ratio determines the potential of the first cluster center, and high values indicate high potential to become a cluster center. The reject rate determines the potential of the first cluster center and those below this value are rejected as cluster centers. In this study, these parameters are adjusted to give the best classification result.

- **Fuzzy C-Mean Clustering**

FCM suggests that each data point belongs to clusters defined by the metrics determined by their degree of membership. The data to be clustered and the number of clusters is the parameters of the function. The function starts with the first guess, but the guess is most likely incorrect. However, the function assigns each data point a membership degree of the set and updates it iteratively. This process moves the cluster centers to the proper location. Mentioned iteration is accomplished by minimizing the distance from a data point to the cluster center, which is valued in accordance with the membership degree of that data point. During reiterations, the function turns the matrix of the final cluster centers, the final MF matrix, and the target values [94]. As a result of these processes, highly accurate and approximate cluster centers are obtained.

FCM separates a collection of n vectors $x_i, i = 1, \dots, n$ into fuzzy sets, and specifies a cluster center for all sets in the way that the objective function of dissimilarity measurement is diminished. $i = 1, \dots, c$ are optionally selected from the n points. The stages of the FCM are, clarified briefly:

First, all cluster's centers $c_i, i = 1, \dots, c$ are chosen in a random way from the n data exemplars $\{x_1, x_2, \dots, x_n\}$.

Secondly, the membership matrix (μ) is calculated with Equation (2.17)

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}, \quad (2.17)$$

where μ_{ij} symbolizes the degree of membership of object j in cluster i . m represents the degree of fuzziness ($m > 1$). d_{ij} is the Euclidean distance between c_i and x_j .

Thirdly, the objective function can be computed with Equation (2.18):

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d_{ij}^2. \quad (2.18)$$

The process continues until it falls below a certain threshold. In the final step, the new c fuzzy cluster centers $c_i, i = 1, \dots, c$ are obtained by Equation

(2.19) [95]:

$$c_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m} \quad (2.19)$$

An example application of the output of FCM is shown in Figure 2.16 [96]. All steps of FCM are applied to divide data into a number of MFs with distinct centers. Every MF is trained by the ANFIS. The parameters (FCM options) in the ANFIS system are as follows and these parameters are set in the interface to give the highest ACC:

- i. The exponent value for the fuzzy partition matrix (U),
- ii. The maximum number of iterations, specified as a positive integer,
- iii. Minimal improvement in objective function between two consecutive iterations, designated as positive scalars.
- iv. Indicator indicating whether to display the objective function value after each iteration.

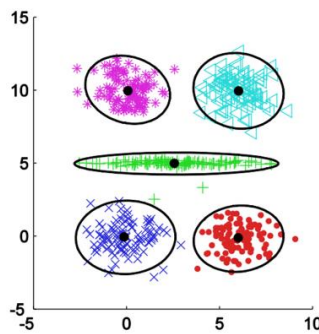


Figure 2.16: An example output of FCM clustering [96]

2.5.3 Deep Learning

Deep Learning is a machine learning method. Like the ANNs in ML, it consists of neurons, just like the human brain. All neurons are interconnected, and this structure creates a layered structure. In networks that basically consist of input, hidden, and output layers, the number of hidden layers on which mathematical calculations are made for deep networks is more than one. That means DL networks are defined as multi-layered (more than 3), deep, and over-complicated transcriptions of the basic ANN in ML. The deep neural network structure in DL is called Convolutional Neural Network.

2.5.3.1 Architecture of CNN

A characteristic CNN architecture mimics the visual cortex. In addition to the input and output layers, it implicates three main hidden layer structures, which are convolutional, pooling, and fully connected (FC) layers. The furthest illustrative features of input images can be extracted via some filters in the convolutional layer

and then selected by the pooling layer. The convolutional layers contain the activation function, which performs certain mathematical processes on it, like ReLU (rectified linear unit). The class can be eventually specified in the FC layer. Whole operations are managed on width, height, and depth for an entry image. In this way, the processing of input is performed in various layers to define ultimate class labels. A visual of a sample study summarizing the CNN model structure is given in Figure 2.17 [97].

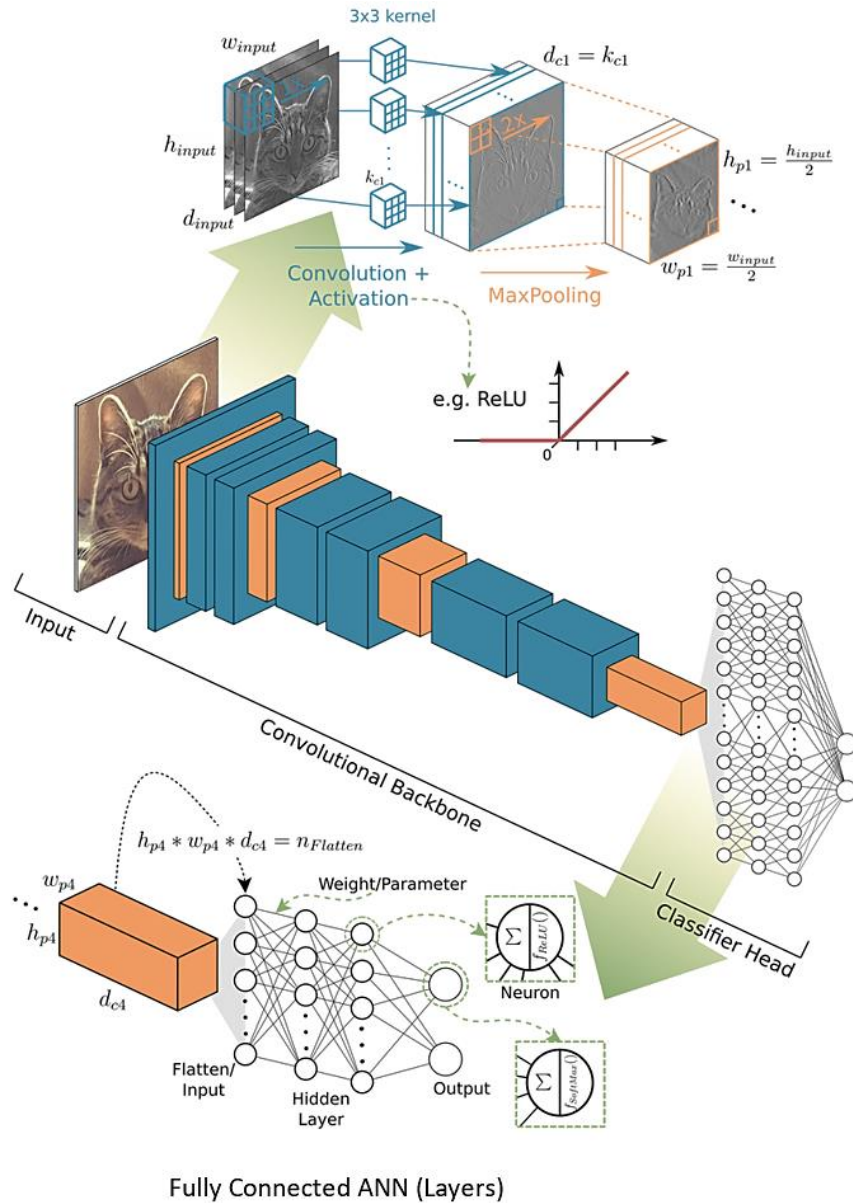


Figure 2.17: An example CNN architecture for image classification [97]

A detailed description of the main layers and other operations, which are dropout, activation function, and batch normalization are presented below.

- Convolution Layer

The convolutional kernel performs by splitting the input image into small segments. The place, where input is taken from in the neuron, is named as receptive field. Filters scan the image and estimate the possibility of its class [98]. In there, a group of filters (with the size of $M \times M$) is used to apply 2D convolution to the image. They slide over the image, and the dot product is obtained between the filter and image parts. An activation function, ReLU, provides the transformation of all negative values to 0 [99]. After the image passes from this layer, a feature map, which gives information about the image, is created. Convolution is represented mathematically as Equation (2.20).

$$f_l^k(p, q) = \sum_c \sum_{x, y} i_c(x, y) \cdot e_l^k(u, v) \quad (2.20)$$

In there, $f_l^k(p, q)$ represents the (p, q) element of the feature matrix. c is the channel index, and l is the layer number. In addition, x and y represent that x^{th} and y^{th} coordinate under consideration of the image. I_c is the input image tensor. $i_c(x, y)$ is a component of I_c , which is component-wise performed multiplication by $e_l^k(u, v)$ sign (or index) of l^{th} layer's k^{th} convolutional kernel k_l . The output of the feature map is $F_l^k = [f_l^k(1,1), \dots, f_l^k(p, q), \dots, f_l^k(P, Q)]$, where P and Q represent the number of rows and columns of feature matrix, respectively [100]. An example of a convolution operation is visualized in Figure 2.18.

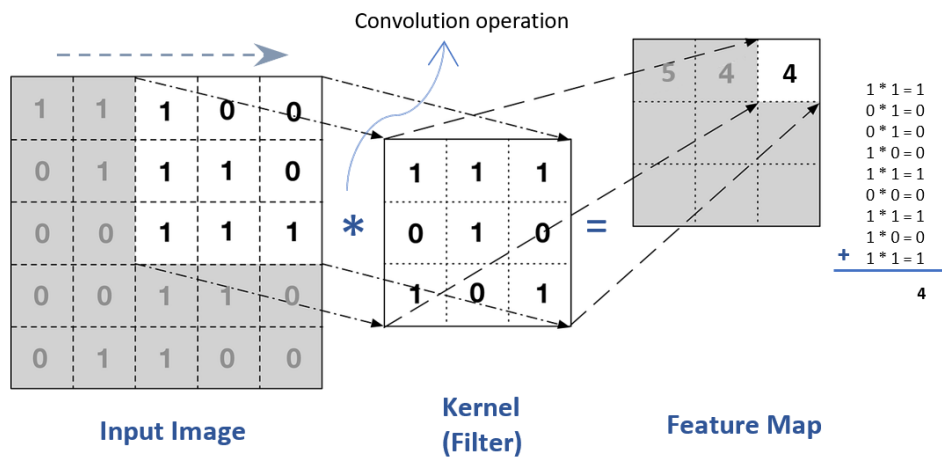


Figure 2.18: Convolution operation

- Pooling Layer

The main function of this layer is to reduce computational costs and discard unnecessary information by performing operations on the convolved feature maps obtained in the convolution layers. In doing so, it reduces the connections between layers. This layer reduces the size of the feature maps as a result. This also provides to decrease overfitting issues. With these operations, there is a bridge function between the Convolutional and FC layers. There are several types such as Max Pooling, which takes the largest value, and Sum Pooling, which takes the sums of the elements [99].

$$Z_l^k = g_p(F_l^k) \tag{2.21}$$

Equation (2.21) shows the pooling operation in which Z_l^k represents l^{th} layer's pooled feature-map for k^{th} input feature-map F_l^k , while $g_p(.)$ is the pooling operation. An example of these two pooling operations is shown in Figure 2.19.

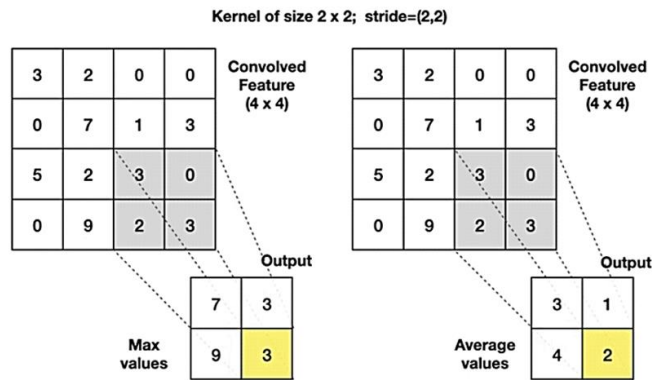


Figure 2.19: Max Pooling and Sum Pooling operations

- Fully Connected Layer

FC layers are usually positioned before the output layer and form the last layers of a CNN. Their main purpose is to connect neurons between two distinct layers. Before the FC layer, the image taken from the layer is flattened and the flatted vector is then fed to the FC layers where mathematical operations are performed. They consist of weights and biases along with neurons. As you pass through these layers, the classification process begins. The reason why connected layers are used here provides better performance when they are connected to each other in binary form.

- Dropout

When features are directly connected to the FC layer, training data may also cause overfitting problems and reduce performance. Therefore, it allows for adding a dropout layer and arranging the network structure. Here, a few neurons are dropped from the networks during training, reducing the size of the model and simplifying it [100].

- Activation Function

An activation function is one of the most significant issues of the CNN model. It is used to learn and make inferences about the sustained and complicated relationship between variables in model networks. For this, it decides where the information will go, that is, it performs the decision of which information should be kept and fire, and which should not. It provides non-linearity of the network. ReLU, Softmax, tanH, and Sigmoid are some of the most preferred activation functions. The activation function for a convolved feature map is specified as follows:

$$T_l^k = g_a(F_l^k) \quad (2.22)$$

where g_a is the activation function, which provides nonlinearity, F_l^k symbolizes convolution output, and T_l^k is the transformed output [100]. An example of the application of ReLU to the matrix is given in Figure 2.20.

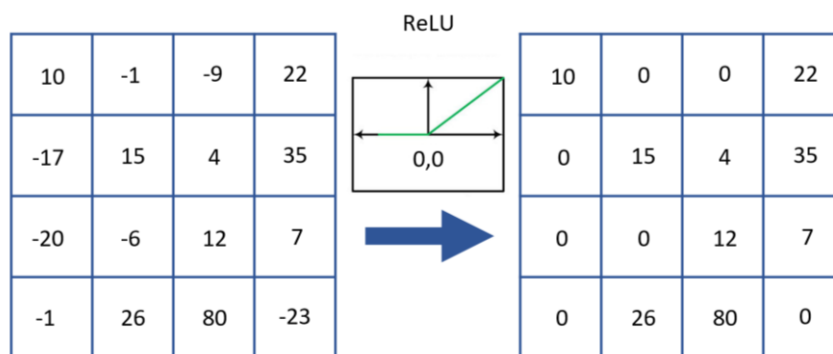


Figure 2.20: An example of application of activation function, ReLU, to matrix

- **Batch Normalization**

It is a normalization method performed among the layers in place of the raw data. It is performed throughout small batches in place of the whole data and provides speed training. It also uses high learning rates that ensure an easy learning process. It is applied to the output of neurons before the activation function applies. The output to be obtained by adding batch normalization to a neuron is formulated in Equation (2.23):

$$N_l^k = \left(\frac{F_l^k - m_z}{s_z} \right) \cdot \gamma + \beta \quad (2.23)$$

where N_l^k is the Batch Normalization output, m_z is the mean of the output of the neuron, s_z is the standard deviation (SD) of the neurons' output. γ and β are learning parameters of normalization. These parameters shift SD and mean, respectively [100].

2.5.3.2 Transfer Learning

Classic CNNs tend to raise computational costs and complexity. They still can be suffering from bias and overfitting issues. To avoid computational costs and complexity issues dropout functions, and batch normalizations can be used as mentioned before. In addition to them, approaches such as residual networks, and Transfer Learning are alternatives to further refine models. TL is based on the approach of utilizing the knowledge learned by pre-trained models in other datasets. It allows to be use of features obtained by pre-trained models, weights, etc. for another classification. Therefore, the CNNs are trained with large datasets like ImageNet which contain a variety of images [101], and become a pre-trained model.

In TL-based applications, the network behaves as a steady feature extractor, when the dataset contains a few numbers of data and possesses a similar to ImageNet. Also, the network ensures to do fine-tuning of the network, if the dataset contains a huge number of data and possesses similar to ImageNet, the latest layer in pre-trained CNN is removed, and the weights are tuned utilizing backpropagation (BP). In another situation, if the dataset contains a large amount of data, and does not possess similar to ImageNet, the training of the model is performed, commencing from the initial layer

or ImageNet’s pre-trained weights [102]. Then, the extracted features are used to be fed to a classifier function such as Softmax for the classification of labels. An example of the TL process from a study [103] is shown in Figure 2.21.

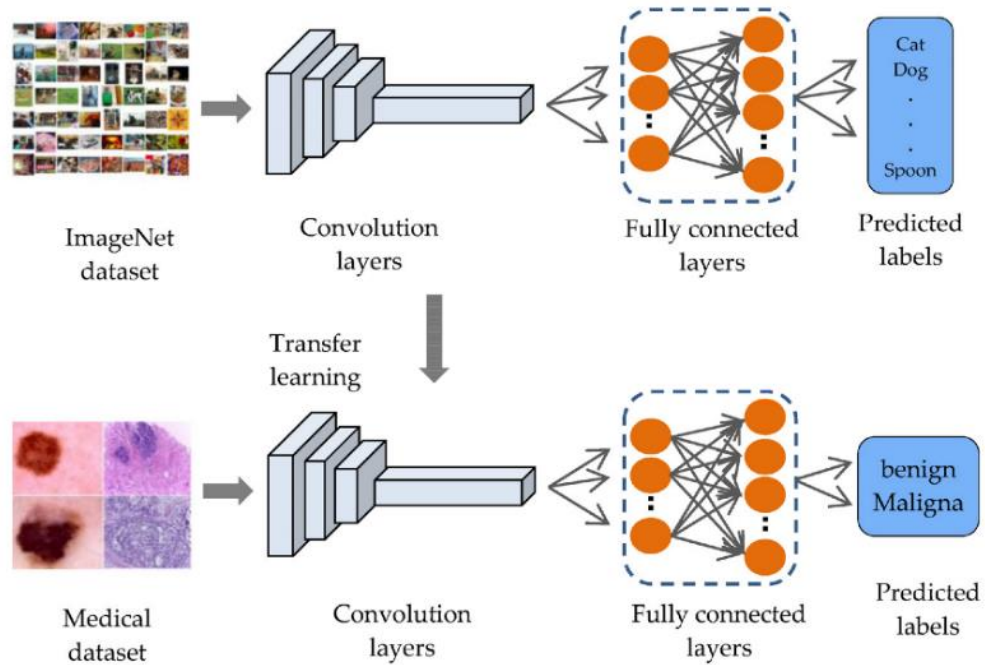


Figure 2.21: An example of medical image classification using TL from a study [103]

In the literature, various CNN models based on the pre-trained approach have been offered as a TL approach. AlexNet and ResNet architectures can be served as one of these models that offer high classification accuracy [30]. Furthermore, it is reported that these models can be beneficial for the TFR image-based classification [68]. The seven well-known and most preferred CNNs, which are AlexNet, SqueezeNet, GoogLeNet, VGG-16, VGG-19, ResNet-34, and ResNet-50, are selected to classify hand gestures [30,68,104]. All of them are pre-trained with the ImageNet dataset. The usage of TF image provides an input naturally consisting of time, frequency, and spatial information of biomedical signal, EMG. Considering these, these architectures’ ability to individuate patterns in IF, and define the non-stationary structures of sEMGs are the motivation for using them in DL-based classification. In these models, only the last layers are rebuilt to maintain their output labels (7×1) for training the HHS images with the seven architectures mentioned above. The other layers are frozen to hold the architectures in the original build. The used CNN architectures and their properties are listed in Table 2.2.

Table 2.2: Characteristics of seven CNNs utilized in this study

Networks	Depth	Parameter ^a	Input Size	Output Size	Description
AlexNet	8	62	227-by-227	7-by-1	Conv:5 – FC:3
SqueezeNet	18	1.3	227-by-227	7-by-1	Conv:18 – FC:0
GoogLeNet	22	7	224-by-224	7-by-1	Conv:21 – FC:1
VGG-16	16	140	224-by-224	7-by-1	Conv:13 – FC:3
VGG-19	19	145	224-by-224	7-by-1	Conv:16 – FC:3
ResNet-34	34	22	224-by-224	7-by-1	Conv:33 – FC:1
ResNet-50	50	26	224-by-224	7-by-1	Conv:49 – FC:3

^a Parameters > 10⁶, **Conv**: Convolutional, **FC**: Fully Connected.

In Table 2.2, the depth, number of parameters, input and output sizes, and description of the layers of the models are given. AlexNet was recommended by Alex Krizhevsky [101]. It is a feedforward CNN and is trained with millions of images of a hundred classes. The first convolutional window in the model starts with the size of (11×11) and decreases to (3×3) as it progresses through the layers. ReLU activation function is used in AlexNet. VGG-16 was recommended by Simonyan and Zisserman. Using small (3×3) convolution filters, VGG-16 is improved and deeper than AlexNet. Like AlexNet, it is trained with millions of images belonging to a thousand classes. The deeper version of VGG-16 is VGG-19, which possesses three more convolutions (weight) layers. VGG architectures use ReLU as an activation function [105].

SqueezeNet is a squeezed (compact) CNN. It has been developed in order to obtain a network model that requires less memory in embedded systems, can be transmitted more easily on the computer, and has fewer parameters. Therefore, it provides less computational cost. Having 1.3 million parameters, it is the model with the least number of parameters compared to the others. It starts with a convolution layer and then eight fire modules come, resulting in a last convolution layer [106]. The

GoogleNet was developed based on the overfitting problem that occurs as the depth increases [107]. This architecture has filters of different sizes. In this respect, the expansion of the networks is carried out rather than the deepening. Three filters, (1×1) , (3×3) , and (5×5) , are applied to the inputs.

ResNet or residual networks are based on residual learning, which can simplify the training of networks [108]. In residual learning, shortcut links bypass one or more layers, allowing inputs to propagate more quickly through these links between layers. The whole type of ResNets, which change depending on layer number, possess their certain residual block. ResNet-34 has 34 layers, and ResNet-50, which is a deeper version of ResNet-34, has 50 layers. However, they possess a distinct residual block plan. In the residual block of the ResNet-34 model, data passes through the (3×3) filter twice, but the ResNet-50 model has bottleneck constructions, where data sequentially passes through filters (1×1) , (3×3) , and (1×1) . The reason for this construction is network deepens. Because of deepness cause increasing the number of weights and computational costs. Thanks to bottleneck construction, the operation's speed rises by reducing the computation load and not losing important information [109]. An example of building block and residual blocks of ResNet-34 and ResNet-50 are drawn and shown in Figure 2.22 based on the study [109].

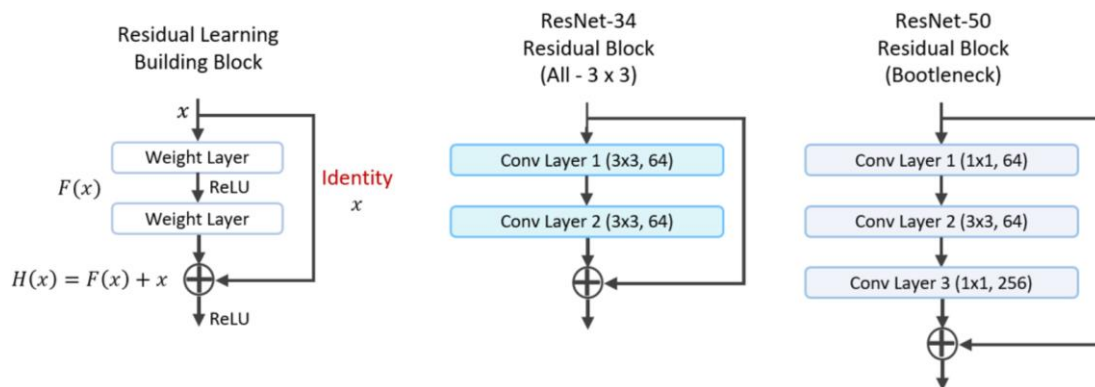


Figure 2.22: Residual blocks of ResNet-34 and ResNet-50

The architecture of all these seven models used in this study is given in Figure 2.23 below. Re-drawings of them were made for this study based on examples of architectures from the study [110] for AlexNet, the study [106] for SqueezeNet, the study [111] for GoogLeNet, the study [110] for VGG-16 and VGG-19, and the study [109] for ResNet-34 and for ResNet-50.

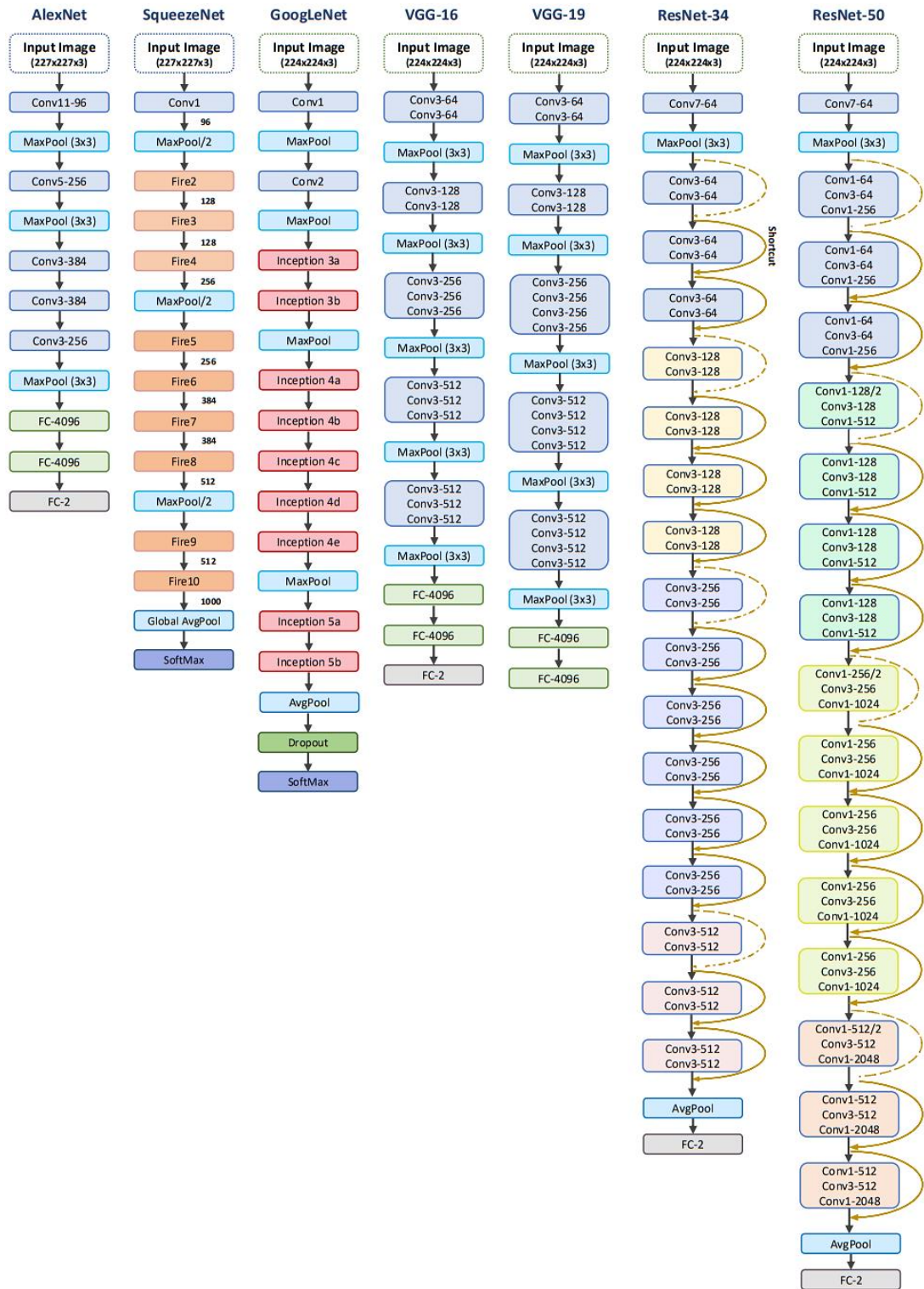


Figure 2.23: The seven CNN architectures used in this study

2.6 Performance Evaluation

The confusion matrix (CM) of each method is obtained from the classification results. Confusion matrices are useful in analyzing the classification process and the resulting outputs. Evaluation metrics can be calculated using these matrices, and it allows comparison by expressing the analysis in numerical values. In the performance evaluation step, several statistical metrics are calculated for each AI approach. Classification accuracy (ACC), and F1-score are calculated for each AI method and their different version applications. The area under the receiver operating characteristic curve (ROC-AUC) is calculated for both DL and ML models. These metrics are shown in the following Equation (2.24) and Equation (2.25):

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.24)$$

$$F_1 - \text{Score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (2.25)$$

FP and TP symbolize false and true positives, and FN and TN symbolize false and true negatives, respectively. The cross-entropy loss function ($\mathcal{H}(P, Q)$), and mean squared error (MSE) are computed for each DL classifier [10,28,30]. The evaluation metrics of the DL architectures are described in the following Equation (2.26) and Equation (2.27):

$$\mathcal{H}(P, Q) = - \sum_a P(a) \cdot \log Q(a) \quad (2.26)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (w_t - \widehat{w}_t)^2 \quad (2.27)$$

where n is the number of data points, and w and \widehat{w} are the actual and predicted values for the point t , respectively. In the cross-entropy loss, P symbolizes objective distribution, Q represents the proximate of objective distribution, and a becomes the number of all classes.

Various metrics have been calculated to evaluate the FL models since they have structural differences compared to other methods. In addition to ACC and F1-Score, sensitivity, specificity, and precision are also calculated according to Equation (2.28), Equation (2.29), and Equation (2.30), respectively. In addition to these, different criteria are also used to evaluate the results of the two FL clustering operations. The coefficient of correlation (R), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) values are calculated for SC and FCM-based methods according to the following Equation (2.31), Equation (2.32), Equation (2.33), and Equation (2.34), respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{GP} + \text{FN}} \quad (2.28)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.29)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.30)$$

$$R = \frac{n \sum v \hat{k} - \sum v \sum \hat{k}}{[[n \sum v^2 - (\sum v)^2][n \sum \hat{k}^2 - (\sum \hat{k})^2]]^{1/2}} \quad (2.31)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{k}_i - \hat{k})^2} \quad (2.32)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{k}_i - \hat{k}| \quad (2.33)$$

$$R^2 = 1 - \frac{\sum (\hat{k}_i - \hat{k})^2}{\sum (\hat{k}_i - \bar{\hat{k}})^2} \quad (2.34)$$

In above equations, the predicted value of \hat{k} is expressed by \hat{k}_i , and the mean value of \hat{k} by $\bar{\hat{k}}$. n is the number of data points.

2.7 Statistical Analysis

In this study, statistical analysis is performed. First statistical analysis is performed for IMF selection. The second analysis is to statistically compare all conducted classification approaches. The classification accuracy of the whole scenarios of FL, ML, and DL models are evaluated separately based on a cross-validation (CV) strategy utilizing a one-way analysis of variance (ANOVA) method with Dunn's posthoc tests. The k-fold CV strategy is used in the ML and FL models, and the stratified k-fold CV (SKCV) strategy is used in the DL models. The k value is taken as 5 for all models.

3 Results

3.1 Results of preprocessing of sEMG Signal

Preprocessing steps were applied in MATLAB[®] to obtain noise-free signals containing movement information from the 490 s long EMG signals obtained from the dataset. sEMG signals applied with a Bandpass filter and Notch filter were taken to the desired frequency range and cleaned from environmental noise. Two different filtered sEMG signals were obtained depending on the frequency range. The filtered 4-channel EMG signals obtained from seven gestures as a result of applying a 5-500 Bandpass filter and 50 Hz Notch filter are shown in Figure 3.1.

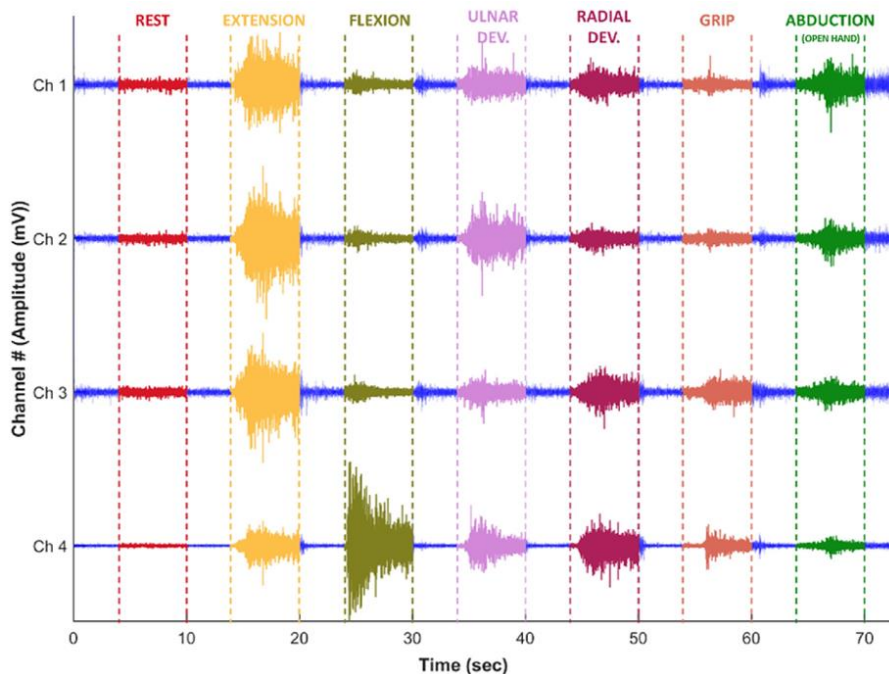


Figure 3.1: The plotting of 4-channel filtered sEMG signal of seven hand gestures

The same segmentation processes were applied to signals with the range of 5-500 Hz and 50-500Hz, separately. Considering the delays that may have occurred during the

realization of the movement, the 1 s at the beginning and the end of the signal were discarded from the 6 s signal, and the signal was taken as a 4 s segment from the steady state moment. Then, the overlapping window was applied to four channels of each movement with rectangular windows of 200 ms and 250 ms in 50 ms increments to increase the number of data and to comply with the time limitation for real-time applications. In this way, a total of 77 (for 200 ms) and 76 (for 250 ms) sEMG segments were obtained from the 4 s segments of a channel. This process was applied to the signals in each channel. For each channel, 80850 segments (30 subjects \times 5 repetitions \times 7 gestures \times 77 segments) for 200 ms long and 79800 segments for 250 ms long (30 subjects \times 5 repetitions \times 7 gestures \times 76 segments) were obtained. Examples of 250 ms-long multi-channel sEMG segments of the Rest state and Extension gesture are shown in Figure 3.2, respectively.

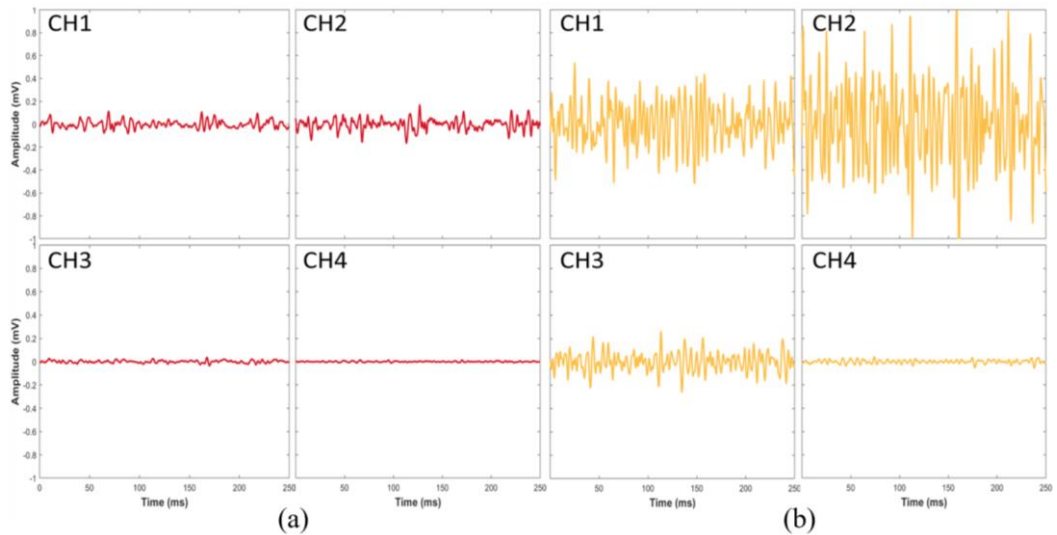


Figure 3.2: 250 ms-long multi-channel sEMG segments of (a) Rest state, and (b) Extension gesture

3.2 Image Generation

All processes at this stage were performed in MATLAB[®]. The segmented four-channel sEMG signals of seven gestures were converted into the TF domain using HHT. In the HHT implementation, IMFs were extracted using the EMD algorithm.

3.2.1 Results of IMF Selection

In order to identify the most distinctive IMFs of the EMD method, two distinct experimental strategies were carried out while generating HHS images. The IMF selection procedure is conducted with a classification experiment that follows a statistical method. In the statistical analyses, the generated various number of IMFs were evaluated whether ones are normally distributed. The statistical analysis results revealed that all of the first three IMFs were statistically normally distributed ($p>0.05$), and almost all of the three IMFs were also normally distributed ($p>0.05$) in the whole dataset. Additionally, the obtained results are in line with the previously reported state-of-the-art studies. It has been reported that the higher frequency components of the sEMG spectrums may contain useful information and may provide a significant contribution to the classification of different gestures [112]. Also, since the top IMFs contain the higher frequency component, they may well represent distinct information when the muscles are steady-state and oscillations are higher. As such, the normally distributed three IMFs were considered for the following analyses.

It can be visually noticed that the oscillations are slowed down in the IMF₃-based HHS and the resolution is dropped at lower oscillations. The first three IMFs found to be statistically more significant were used to generate HHS images using the f_s as 2000 Hz. For ML and FL implementation, the obtained HHS images of 4 channels were used separately in the feature extraction stage ignoring titles, and axis labels. According to the IMF selection results, HHS images created using the first three IMFs obtained from 200 ms long segments were used for ML and FL studies. An example of these images is shown in Figure 3.3.

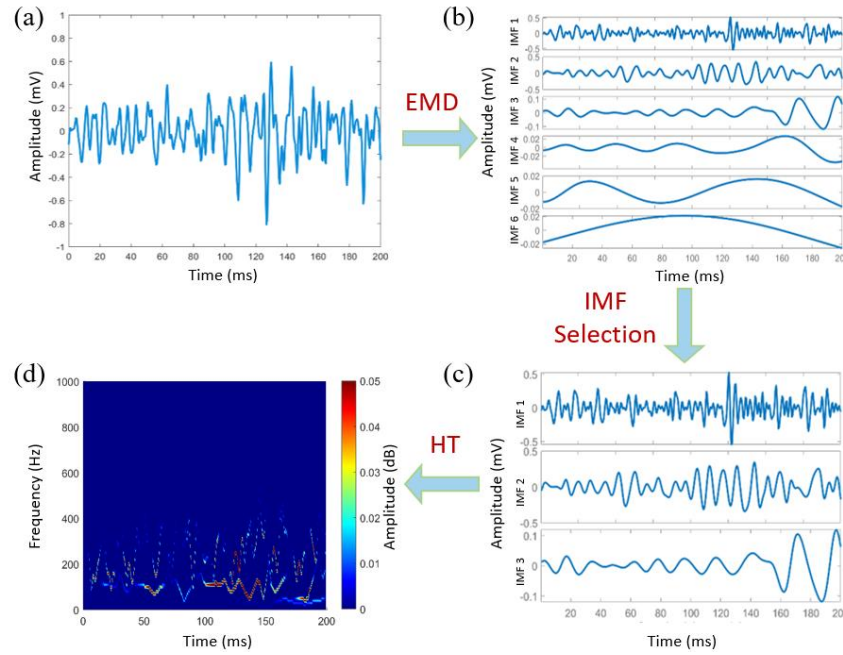


Figure 3.3: The whole process of obtaining HHS images using first three IMFs of EMG signal: (a) 200 ms sEMG signal, (b) first six IMFs, (c) selected first three IMFs, and (d) HHS image using first three IMFs

The generated 3D HHS using the first three IMFs can be represented as a volume or a 2D image using a color mapping. The processes of generation of 2D color-mapped images of 200 ms-long segments are shown in Figure 3.3. Hence, the instantaneous energy of sEMG segments becomes a color representation in the 2D image. The color scale of all TF images is normalized to create a fair classification.

After the normally distributed IMFs were selected, their different combinations are generated as IMF_1 , IMF_2 , IMF_3 , IMF_{1-2} , IMF_{1-3} , and IMF_{1-to-3} . To be more precise, IMF_{1-3} stands for HHS formed by using IMF_1 and IMF_3 , and IMF_{1-to-3} stands for HHS formed by using IMF_1 , IMF_2 , and IMF_3 . These different combinations were used in DL models. The 2D HHS images generated using IMF_1 and IMF_3 with 250 ms-long segments are shown in Figure 3.4-(a) and Figure 3.4-(b), respectively.

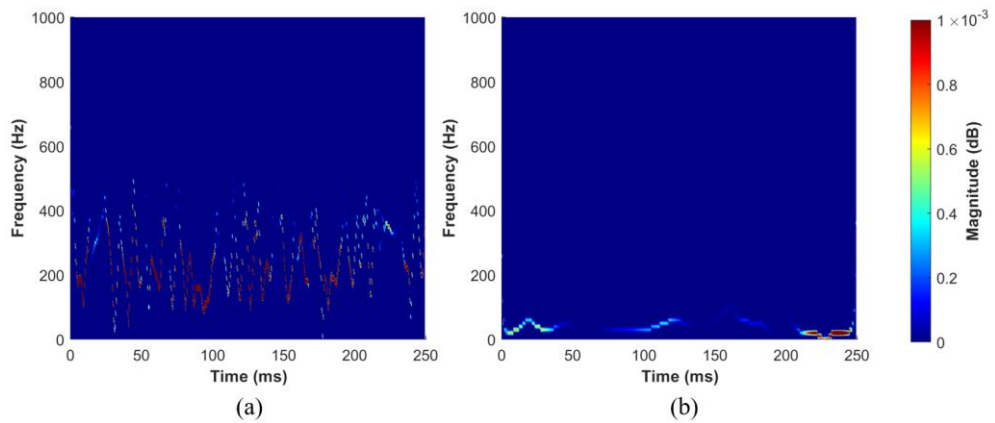


Figure 3.4: HHS images generated using (a) IMF₁ and (b) IMF₃ with 250 ms-long segments

3.2.2 Results of Different Length HHS Images

For the DL application, TF image acquisition was applied for 200 ms and 250 ms signals separately. 200 ms and 250 ms HHS images of the extension movement are given in Figure 3.5. In DL, 200 ms and 250 ms lengths were compared according to different IMF combinations.

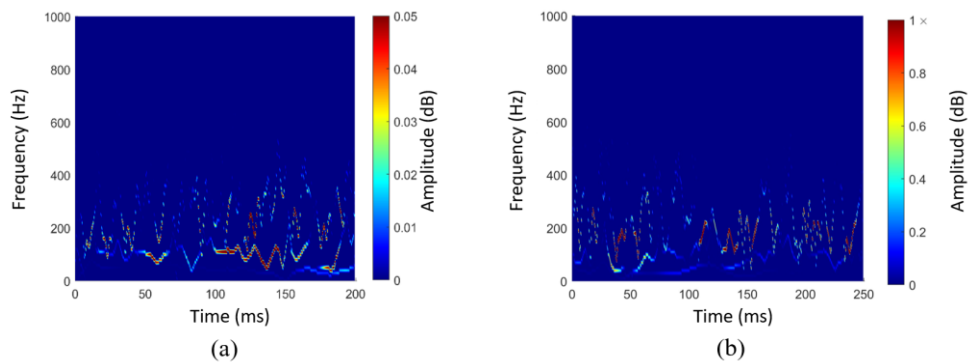


Figure 3.5: HHS images of the extension movement using first three IMFs of (a) 200 ms, and (b) 250 ms segments

3.2.3 Results of TF Image Fusion

For DL, HHS generated using 4-channel sEMG segments were combined into a series of images, ignoring titles, and axis labels. Thus, while the temporal and spectral information provided by each channel is preserved separately because of the advantages of TFA methods, the jointly spatial information is also preserved by multiple channels represented as a fused image. The channel placement in the fused image represents the spatial information, while the color information is the expression

of the spectral energy value of the sEMG signals. Besides, there is a requirement for joint coordination of information of several muscles in order to distinguish the corresponding gesture from sEMG signals because the multi-channel sEMG signals are mutually relevant. Thus, the generated multi-channel combined spectral image may offer different visualization for each gesture. Also, this representation may enable the extraction of implicit correlations between the multi-channel sEMG signals. Hence, it may enhance the ability of feature expression of EMG signals because valuable spatial information is preserved. Moreover, if additional EMG channels are provided, the proposed method can be easily evolved by adding TF images of additional channels to the main fused spectral image. Eventually, each class includes an equal number of images resulting in a balanced dataset, and the separate 4-channel images and the fused TF images have modeled the output with corresponding gestures. Therefore, the train and test datasets are channel-wise scaled to zero-mean and unit variance. An example fused image created by combining the HHS images of four channels is shown in Figure 3.6. The dotted white stripes and channel names here are used to describe the image [62].

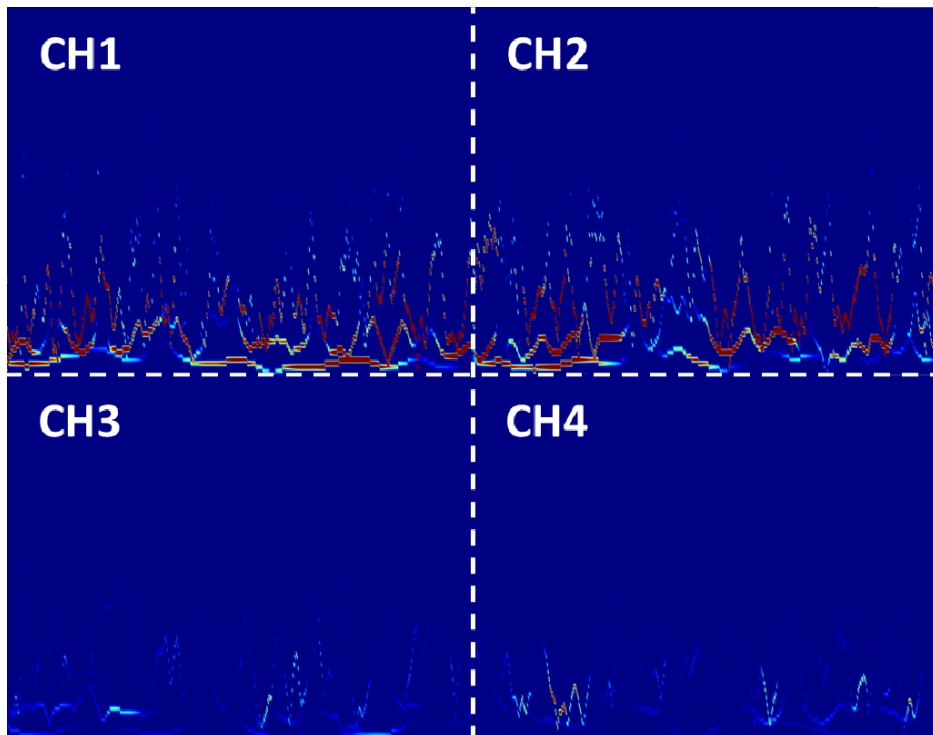


Figure 3.6: Fusion image of four channel

3.2.4 Results of HHS Images of Different Gestures

The fused images given as input to the deep networks are the contiguous aligned versions of the clean TF images. Significantly different HHS images were obtained for each movement. As an example, the rest and extension movement are visualized here. 3D HHS representations obtained from 250 ms-long sEMG segments of Rest state and Extension gestures using the first-three IMFs are demonstrated in Figure 3.7-(a) and Figure 3.7-(b), and the 2D color-mapped images of corresponding HHS are demonstrated in Figure 3.7-(c) and Figure 3.7-(d), respectively. In those 2D TF images, the x and y axes represent time and frequency, and the z-axis represents the instantaneous energy of sEMG segments in the 3D HHS. It can easily be noticed that there is a significant visual difference between the two gestures in the images.

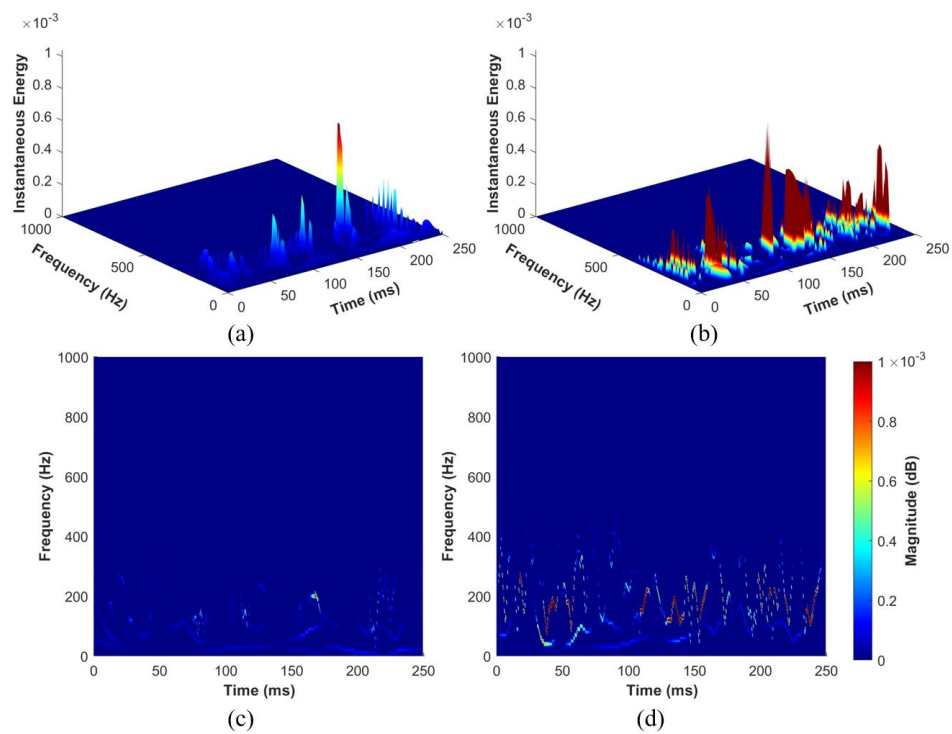


Figure 3.7: 3D HHS representations: (a) 250 ms-long sEMG segments of Rest state and (b) 250 ms-long sEMG segments of Extension gesture using the first-three IMFs. (c) 2D color-mapped images of corresponding HHS of Rest, and (d) 2D color-mapped images of corresponding HHS of Extension

3.2.5 Results of Snapshot Images

In addition, in ML application, snapshot images of 200 ms EMG signal and first three IMFs were taken to be used in comparison with TF images' performance. As an example of them, the images in Figure 3.8-(a) and Figure 3.8-(b) can be given. Totally, 80850 snapshots were obtained from EMG signal segments and the first three IMFs, and HHS, separately [65].

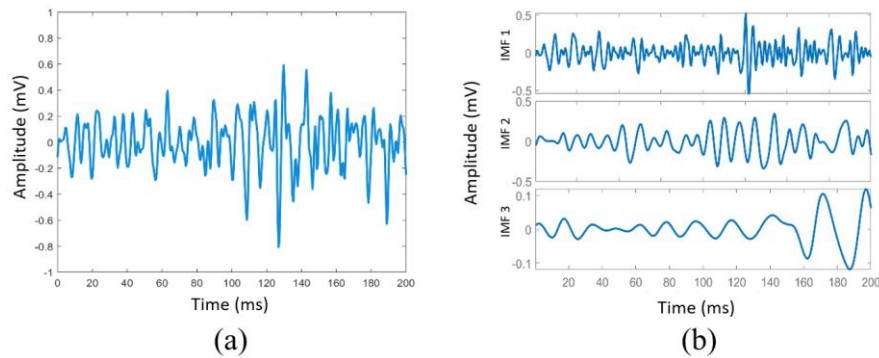


Figure 3.8: Snapshot image examples: (a) EMG, and (b) IMFs

3.3 Feature Matrix Generation

As a result of applying HHT to 200 ms EMG segments, HHSs were obtained and six features of the spectra were extracted with MATLAB[®], considering the principle of image-based feature extraction from these spectra. At this stage, colored HHS images were first converted to grayscale, and then GLCM was calculated. Six features expressing different statistical measures of GLCMs were calculated for each AI method. With the dataset created using six features extracted from the spectrum images, a new information representation using the time series analysis method was obtained. An 80850×24 dataset was created, with the twenty-four columns (6 features x 4 channels) being the feature (input) column.

In ML and FL applications, mean, variance, energy, and entropy features were common, but two of the six features were different. While skewness and kurtosis were used as the fifth and sixth features in ML, contrast and homogeneity were used in FL. In addition, different strategies were tried on the dataset. For the ML model, each feature was represented channel-wise. In addition, all these GLCM feature extraction

processes are repeated for snapshot images of EMG and IMFs for use in ML, and a feature matrix of the same dimensions was obtained. For FL, each feature value obtained from the four channels was then averaged for data reduction purposes and the number of columns was reduced to 6. Feature matrices were normalized before being used in ML and FL. These data were separated as 80% (64680) for training and 20% (16170) for testing.

3.4 Results of ML-based Classification

After the implementation of dimensionality reduction, five features are determined by the PCA except for the mean. The 5-fold CV technique was used to prevent overfitting. It provided an accuracy value of over 83.20% in 28 of 29 ML models. An accuracy value of 60.80% was obtained only in the Coarse Tree model. The average of the accuracy values obtained from all models is over 86%. The five highest accuracy values in Table 3.1 were obtained with Cubic SVM (90.87%), Quadratic SVM and Medium Gaussian SVM (89.70%), Ensemble Bagged Trees (89.60%), and Ensemble Subspace kNN (89.40%), respectively. When the models were examined in terms of type, DT-based models achieved the lowest classification performance (approximately 77.00% on average). The highest classification performance was achieved by SVM-based models with an average value of 89.20%. After SVMs, NNs and discriminant analysis-based models were found generally more successful (ACC of >87.00%).

Table 3.1: The accuracy values of highest five ML models for HHT-based features

Method	Model	ACC (%)
HHT-based GLCM Features	Cubic SVM	90.87
	Quadratic SVM	89.70
	Medium Gaussian SVM	89.70
	Ensemble Bagged Trees	89.60
	Ensemble Subspace kNN	89.40

The best performance results for the three ML approaches were given in Table 3.2. The EMG and IMF snapshot-based approaches were behind the HHT-based approach in all tested ML methods. In EMG signal-based approach, the best validation ACC was obtained as 71.34% by Random Forest. The F1-score was 70.89% and the mean

AUC was 0.78. In the IMF-based approach, the best validation ACC was obtained as 82.13% by Bagged Trees with five features. The F1-score was 82.74% and the mean AUC was 0.90. Here, the feature extraction approach by using the first-three IMFs increased 10.79% of the classification performance compared to the signal-based approach. The GLCM-based HHS features achieved the highest ACC 90.87% with the Cubic SVM method. The ACC values of punch, flexion, open hand, and ulnar deviation were obtained above 92.00%, whereas other gestures have ACC values above 80.00%. The average F1-score was calculated above 90.84%. When the AUC values were examined, the mean AUC was ~ 1 for the Cubic SVM. The precision, recall, and specificity of the model were obtained as 91.14%, 90.87%, and 98.48%, respectively [65].

Table 3.2: Classification results of the three different approaches for ML

Method	Model	ACC (%)	F1-Score (%)	AUC
EMG Snapshots	Random Forests	71.34	70.89	0.78
IMF Snapshots	Bagged Trees	82.13	82.74	0.90
HHS Images	Cubic SVM	90.87	90.84	~ 1

In addition, an error analysis between gestures for HHS-based method was performed using the confusion matrix in Figure 3.9.

Actual Labels	Extension	1869	201	20	3	145	15	57
	Flexion	7	2204	25	35	14	3	22
	Open Hand	11	4	2156	51	37	4	47
	Punch	12	7	13	2245	2	9	22
	Radial D.	16	18	59	65	1996	6	150
	Rest	57	123	2	18	28	2077	5
	Ulnar D.	17	43	37	22	39	5	2147
		Predicted Labels	Extension	Flexion	Open Hand	Punch	Radial D.	Rest

Figure 3.9: Confusion Matrix of HHS-based classification for ML

3.5 Results of FL-based Classification

The dataset was prepared to be trained with the k-fold CV method in ANFIS architecture and the k value was chosen as 5. Before classification, SC and FCM clustering techniques were applied to the new dataset consisting of six features of the images, respectively. Hybrid optimization was applied for both methods, and the epoch number was accepted as 100 and the error tolerance value as 0 during the training process. In FCM, the parameters calculated using the '*genfis*' command were transferred to the ANFIS interface. Since cross-validation was performed in both clustering methods, the parameters that gave the best results for each repetition and the MF numbers were determined and applied to the data.

The total data (80850×24) was divided into five equal pieces (P1, P2, P3, P4, P5) with 16170 lines of data in each part. Which parts are used as training and test data, MF numbers (number of rules) obtained as a result of the entered clustering parameters and providing the best classification result are shown in Table 3.3. For SC, the values of the range of influence, squash factor, accept ratio, and reject ratio parameters are given in Table 3.3, with their original order in the ANFIS interface. For FCM, the fourth parameter (iv) of it takes the value of true or false, and in this study, the true or false status of this parameter is shown with '0' and '1'. By considering the range of values that other FCM parameters can take, the parameters that are most suitable for the dataset and give the best results were determined and applied for each repetition [61]. These parameters are listed as (i, ii, iii, iv), respectively, Table 3.3. Membership functions change depending on parameters and pieces of total data.

Table 3.3: Applied parameters for SC and FCM methods

Method	Repetition	Training Sets	Test Set	Parameters	MF Number
SC	1	P2, P3, P4, P5	P1	0.24 / 1.3 / 0.4 / 0.395	30
	2	P1, P3, P4, P5	P2	0.26 / 1.0 / 0.4 / 0.395	38
	3	P1, P2, P4, P5	P3	0.24 / 1.4 / 0.4 / 0.395	25
	4	P1, P2, P3, P5	P4	0.3 / 1.33 / 0.4 / 0.395	16
	5	P1, P2, P3, P4	P5	0.3 / 1.33 / 0.4 / 0.395	19
FCM	1	P2, P3, P4, P5	P1	1.5 / 85 / 0.001 / 0	18
	2	P1, P3, P4, P5	P2	1.5 / 85 / 0.001 / 0	16
	3	P1, P2, P4, P5	P3	1.5 / 85 / 0.001 / 0	26
	4	P1, P2, P3, P5	P4	1.5 / 85 / 0.001 / 0	17
	5	P1, P2, P3, P4	P5	1.5 / 85 / 0.001 / 0	16

The ANFIS network structure with 19 rules (with MF) created in MATLAB[®] for one of the best repetitions in the CV is shown in Figure 3.10 as an example.

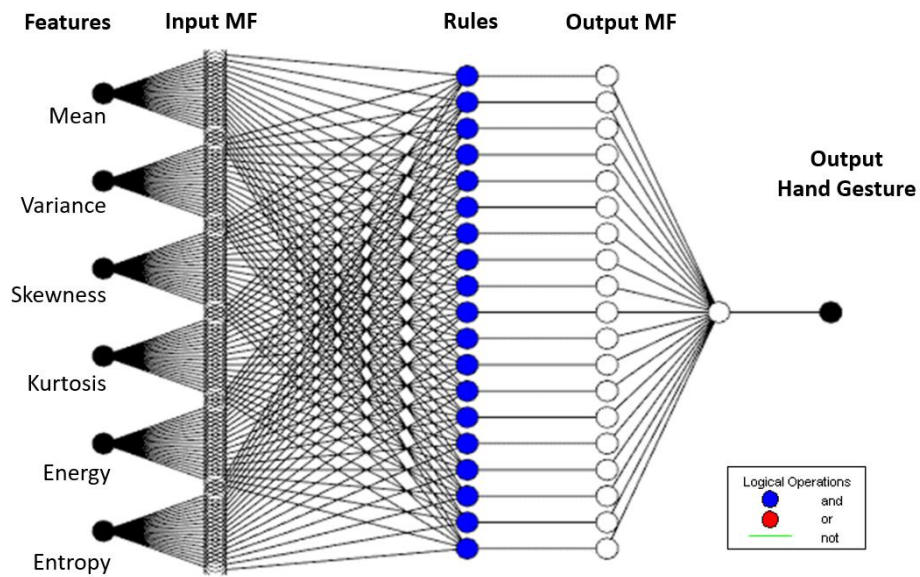


Figure 3.10: An example of ANFIS model with 19 rules resulting from classification Confusion matrices of each repetition were obtained from the classification results. Classification results were visualized in matrix form as a true class and predicted class. The confusion matrices of the first iterations of the test data of the models are shown in Figure 3.11 for SC and in Figure 3.12 for FCM. Within the 16170-test data, there

are 2310 data for each movement. Sensitivity, specificity, precision, accuracy, and F1-Score metrics were calculated for each repetition of each clustering method using the values in the confusion matrices, and their averages are shown in Table 3.4.

Actual Labels	Extension	2221		2	2	23	21	41
	Flexion	1	2214	56	23	4	5	7
	Open Hand	16	3	2046	121	23	98	3
	Punch	3	3	2	2272	20		10
	Radial D.	17	2	10	50	2197	22	12
	Rest	12	3	2	3	172	2076	42
	Ulnar D.	1	5	7	4	3	3	2287
		Extension	Flexion	Open Hand	Punch	Radial D.	Rest	Ulnar D.
Predicted Labels								

Figure 3.11: Confusion matrix of first iterations of the test data for SC

Actual Labels	Extension	2186	2	8	6	34	18	56
	Flexion	7	2186	72	18	3	10	14
	Open Hand	13	11	1913	182	32	157	2
	Punch	24	2	1	2217	21	12	33
	Radial D.	32	4	15	52	2155	25	27
	Rest	36	5	15	21	146	2039	48
	Ulnar D.	19	8	11	6	17	6	2243
		Extension	Flexion	Open Hand	Punch	Radial D.	Rest	Ulnar D.
Predicted Labels								

Figure 3.12: Confusion matrix of first iterations of the test data for FCM

Table 3.4: The results of evaluation metrics for each repetition of clustering methods

Method	Repetition	Sensitivity (%)	Specificity (%)	Precision (%)	ACC (%)	F1-Score (%)
SC	1	94.70	99.12	94.81	94.70	94.69
	2	95.18	94.07	94.00	94.62	94.59
	3	94.52	92.90	92.76	93.69	93.63
	4	94.93	92.76	92.58	93.82	93.74
	5	92.58	90.53	90.29	91.53	91.42
	Average	93.61	93.67	92.89	93.88	94.38
FCM	1	92.39	98.73	92.50	92.39	92.35
	2	92.45	89.96	89.80	91.18	91.11
	3	92.38	91.63	91.53	92.00	91.95
	4	93.22	90.42	90.11	91.77	91.64
	5	91.99	89.77	89.49	90.85	90.72
	Average	91.55	91.64	90.69	92.10	92.49

Table 3.4 presents an evaluation on the basis of repetition. When all metrics are examined, it is seen that SC offers better performance. Its average values of sensitivity, specificity, and precision were obtained as 93.61%, 93.67%, and 92.89%, respectively. In addition, the averages of ACC and F1-score were calculated as 93.88% and 94.38%, respectively. However, the averages of ACC and F1-score for FCM were calculated as 92.10% and 92.49%, respectively. It is seen that there is a difference of almost ~2% between the average metric values of the FCM and the SC values.

When an evaluation is made on the basis of movement, classification values for each hand movement have been obtained differently from each other, as seen in the confusion matrices. The matrices in Figure 3.11 and Figure 3.12 belong only to the first iteration, and the average values calculated from the test data results of the SC-ANFIS and FCM-ANFIS methods were examined for each movement when the matrices obtained in the other iterations were included. The highest accuracy values were obtained with punch movement as 99.00% for SC-ANFIS and 97.10% for FCM-ANFIS. After punch, flexion movement achieved over 95.00% accuracy (98.36% for SC and 95.97% for FCM). According to the performance values, open hand, resting,

and radial deviation movements, respectively, provided over 93.00% accuracy for both clustering methods. Finally, the movements that offered the least classification accuracy in both clustering methods were ulnar deviation (89.87% for SC and 88.57% for FCM) and extension (88.27% for SC and 82.81% for FCM), respectively [61]. In order to evaluate the performance of both methods (SC-ANFIS and FCM-ANFIS), R , RMSE, MAE, and R^2 values calculated separately for training and test data are presented in Table 3.5.

Table 3.5: The results of R , RMSE, MAE, and R^2 values for training and test data

Model	R	RMSE	MAE	R^2
<i>SC (Training)</i>	0.9919	0.8318	0.1937	0.9841
<i>SC (Test)</i>	0.9624	0.8973	0.2671	0.9327
<i>FCM (Training)</i>	0.9828	0.8796	0.3296	0.9658
<i>FCM (Test)</i>	0.9554	0.9098	0.3858	0.9136

3.6 Results of DL-based Classifications

The TF images of Hilbert-Huang spectra were resized to the input dimension of each TL-based architecture with a high DPI resolution. Each TF image was generated in its natural shape with a blue background to represent the lowest energy levels, and devoid of axis labels, title, and colorbar to ignore unnecessary pixels that do not carry information about the sEMG signals. The parameters for HHS-based architectures are described in the previous section. SKCV is adopted to minimize sampling bias and to evaluate the robustness of the proposed models [30]. SKCV was used to provide population-based inference and the generalization for the trained model. The SKCV strategy was utilized based on a subject-correlated approach where the trained model contains a mixed as well as an equal amount of data from all subjects.

In the training phase, the pre-trained architectures with the ImageNet dataset are re-trained with mentioned CV strategy. The fold value k was taken as 5. In this CV, the whole TF image dataset is indiscriminately allocated as training and testing image subsets. Therefore, 80% of the total images were utilized to train and 20% of images were utilized to test the models in each fold. It should be noted that each fold has an

equal number of training, validation, and test data from each subject as well as each class in the SKCV. Thus, each fold represents the entire dataset without overlapping and the test data is never used in the training phase. In order to prevent bias, data is randomly shuffled before the CV strategy is processed. CV strategy revealed the results were robustly evaluated and the trained models were not overfitted.

During the training phase, all layers of the pre-trained architectures were frozen and exactly preserved, except the output layer, which was modified as 7x1 to classify 7 different gestures. Additionally, the hyperparameters were determined with manual tuning based on several empirical trials. Adam Optimizer was used with a batch size of 64, and 10^{-3} learning rate. The epoch value has been set at 50 to observe classification performance in all architectures under the same conditions. Finally, the Softmax function is used to normalize the outputs. All procedures were conducted on an Nvidia GeForce RTX 2080 TI GPU with 64 GB RAM using Keras libraries with backend TensorFlow 2.8, cuDNN 8.1, and Cuda 11.2.

3.6.1 Classification Results of IMF and WL Combination

HHS images generated from mentioned IMF combinations were empirically tested to choose the best combination. In the preliminary experiments, 250 ms-long sEMG segment length and ResNet-50 architecture were selected because while 250 ms window length (WL) was reported as the top limit value for online tasks and yielded better performance among the other WLs [67], ResNet-50 achieved better classification performance among the other deep networks [68]. Additionally, the effects of varieties of CNN architecture on classification performance were also examined in the following experiments. The comparison of HHS method test results of 12 different classification models obtained by using ResNet-50 architecture is shown in Figure 3.13.

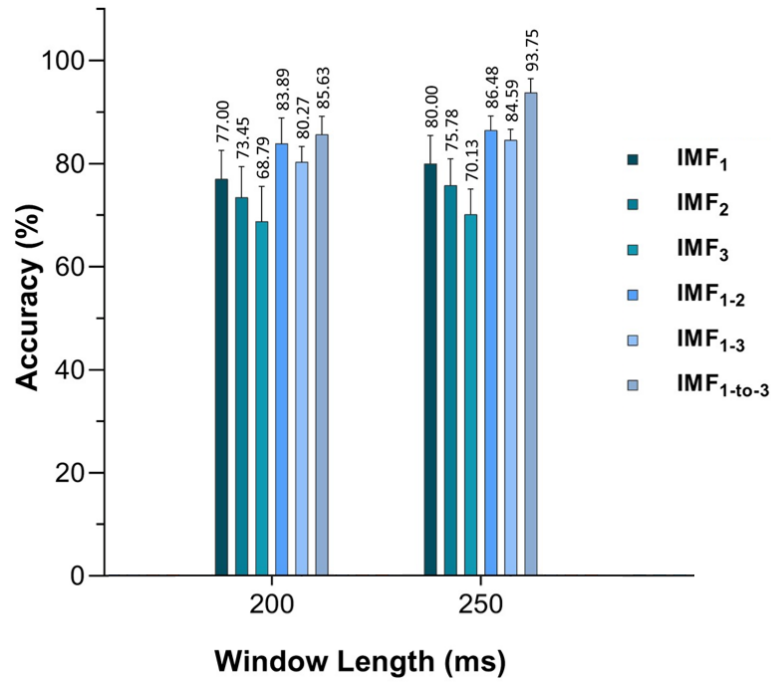


Figure 3.13: The test results of different IMF combinations using ResNet-50 model

Here, the results are obtained for different combinations with the SKCV strategy and it was also examined the WL variation effect on the classification performance to select the best WL-IMF combination. In line with this aim, sEMG segments (200 and 250 ms) were considered, then HHT was applied to these segments to generate HHS images. Accordingly, overall results revealed that the best ACC values were achieved by IMF_{1-to-3}, which achieved the best ACC for 250 ms. Among all IMF and WL combinations, IMF_{1-to-3} achieved the highest average ACC of 93.75% with 250 ms length, while IMF₃ yielded the lower ACC in all cases. The worst ACC was 68.79% in 200 ms WL with IMF₃. According to obtained results, the ACC values increased as the WL increased, regardless of the changing IMF combinations in all cases. However, considering that WLs under 250 ms can be used in real-time applications [67], 250 ms is the border of this and can be used for classification. Due to the superior performance provided by IMF_{1-to-3} and 250 s-long sEMG segments than other real-time WL limits, this combination is used in the rest of the experiments. The increase in ACC at larger WLs may be associated with increased resolution, which is correlated with changing sample size.

3.6.2 Classification Results of CNN Models

Finally, in order to choose the best CNN model in the classification of gestures, 7 different CNN architectures were re-trained with the TL method. The average SKCV results obtained using 250 ms-long sEMG segments and IMF₁₋₁₀₋₃ are presented in Table 3.6. In feeding the TL models with HHS using SKCV, the average AUC values are yielded in the range of 0.84-0.97. Considering the average validation accuracy, only four architectures; GoogLeNet, VGG-19, ResNet-34, and ResNet-50, yielded over 90.00% accuracy. Best performances are obtained by two ResNet architectures. While ResNet-34 achieved ~2% better validation accuracy (92.23%) than VGG-19 and GoogLeNet, it still falls behind ResNet-50. Notice that the best average training ACC and validation ACC are 96.12% and 93.75±1.71%, respectively, by using the pre-trained ResNet-50 model. Also in this model, the training loss is 0.023 and the validation loss is 0.044. The AUC is 0.97, the F1-score is 93.98%, and MSE is 0.321. In the statistical tests, while no significance ($p=0.378$) was found between ResNet models, both outperformed ($p<0.05$) other architectures [62].

Table 3.6: Classification results of seven CNN architectures using first three IMFs of 250-ms sEMG segments

Network	^t ACC	^v ACC	^t Loss	^v Loss	MSE	F1	AUC
AlexNet	82.31	81.56	0.215	0.255	0.587	82.59	0.86
SqueezeNet	81.93	80.73	0.195	0.259	0.576	79.33	0.84
GoogLeNet	91.34	90.36	0.131	0.117	0.521	90.77	0.92
VGG-16	86.77	85.38	0.146	0.195	0.541	85.41	0.89
VGG-19	92.56	90.65	0.098	0.136	0.484	90.15	0.92
ResNet-34	96.05	92.23	0.046	0.089	0.356	93.12	0.95
ResNet-50	96.12	93.75	0.023	0.044	0.321	93.98	0.97

*Loss: Cross-Entropy Loss, all ACC, PRE, and F1-Score values are given as percentage.

^t: Training and ^v: Validation phase.

4 Discussion

Three different artificial intelligence approaches were compared with other methods within themselves.

4.1 Evaluation of ML-based Classification Results

According to the results in Table 3.1, the three best classifications were obtained with SVM-based models. The SVM method offered better results than other ML models. Although there were misclassified gestures, the margin in the SVM model better was able to determine the boundaries between classes. In addition, a large dataset was created with the HHS-based features obtained by data transformation in this study, and the advantage of SVM to work with large datasets was effective in obtaining good results here. Also, it does all this with less memory requirement. In addition to all these, it has the capacity for generalization and fights the overfitting issue better [113]. Generally, it is compared to ANN, but as in this example, SVM models gave better results.

As the performance metrics of all the approaches given in Table 3.2 are compared, it is noticed that the TFR-based approach outperforms the time-domain EMG signal and IMF-based approaches [114]. Besides, using the GLCM of TFR for feature extraction has been shown to be an effective approach for hand gesture classification with sEMG signals. The GLCM-based HHS features preserve the information acquired from HHT that determines the intrinsic features of short EMG signals, signal frequency characteristics, and also the changing of FD characteristics over time [115].

When the classification performances of the movements are examined, the misclassification rate is higher between extension and flexion, radial deviation and

ulnar deviation, extension, and radial deviation, and rest and flexion than in other combinations. The reason for this may be related to the similar contraction while performing the different gestures [3]. For instance, when the hand is during the radial deviation, the hand may sometimes be in a form similar to the extension due to the anatomical structure. This may have caused a similar contraction in the muscles. Hence, the distinguishment between HHS of these gestures may reduce due to this issue. Similarly, tend of EMG activity of radial and ulnar deviation was approximated, and muscle activation was similar in the related channels and amplitude levels. This may have caused the misclassification of these gestures. Rather than this misclassification, the overall performance is remarkable.

4.1.1 Comparison with State-of-the-Art Studies

To assess the effectiveness of HHS-based image features on classification performance, it is summarized the performance results of some recent studies published in the literature in Table 4.1. Huang et al. [116] applied a binary tree (BT) based SVM method to classify 13 movements. They collected the data from more channels and fewer subjects compared to the presented study. Despite using 10-fold cross-validation, they achieved a 2.67% lower ACC than this study's best model. Similarly, Rabin et al. [38] obtained 13.57% less ACC with STFT-based methods. A multi-channel study by Sattar et al. [117], which performed 10-fold cross-validation and classified fewer movements, obtained ~0.17% less ACC than the HHS-based method with half the number of subjects. Also, Kukker et al. [118] used HHT and ANN for movement recognition but, they used IMF₁-based TD features and yielded an average ACC of 86.20%. Similarly, the study of Benalcázar et al. [119], which used the dynamic time warping algorithm and more channels obtained low ACC compared to this study.

Some studies have reported more successful results on the classification. Devaraj and Nair [39] and Mahmood et al. [120] used TD features in the kNN method and achieved a classification ACC of 93.00% and 98.90% ACC, respectively. Qi et al. [121] used a combination of TD and FD features and reached 95.10% ACC with the GRNN method. In the study of Laksono et al. [122], the combination of Teager–Kaiser energy operator (TKEO) and kNN was used for EMG data classification with TD features.

Some of these reported studies suffer from some limitations even though they have reported remarkable results. Some of them used an unbalanced dataset, others datasets suffer from subject-biased, and some of them consider only a few gestures. As a result, it has been concluded that the proposed method may be an effective alternative approach for feature extraction from sEMG signals in hand gesture recognition.

Table 4.1: Comparison Table of ML Studies

Study	Year	Mov.	Ch.	Subj.	Model	Method	Val.	ACC (%)	F1 (%)
[118]	2016	6	2	10	ANN	IMF ₁ -based features	TTS	86.20	N/A
[116]	2017	13	16	8	BT-SVM	TD features	10-fold	88.20	N/A
[121]	2019	9	16	1	GRNN	TD and FD features	TTS	95.10	N/A
[39]	2020	7	8	32	kNN	TD features	TTS	93.00	N/A
[38]	2020	6	2	5	kNN	STFT-based	10-fold	77.30	N/A
[120]	2021	18	8	3	Fine kNN	TD features	TTS	98.90	N/A
[122]	2021	3	3	5	TKEO+Subspace kNN	TD features	TTS	96.67	N/A
[117]	2021	4	8	15	kNN	TD features	10-fold	90.70	N/A
This Study	2022	7	4	30	Cubic SVM	HHT features	5-fold	90.87	90.84

***Mov.:** Movement, **Ch.:** Channel, **Subj.:** Subject, **Val.:** Validation, **TTS:** Train-Test Split.

4.2 Evaluation of FL-based Classification Results

When the results obtained were evaluated, better results were obtained with the SC-based method for each movement. When comparing the seven hand gestures, the best classified gesture is the punch gesture. However, the success of other hand movements is not much different from the success of it. Considering the high level of muscle activation in punch movement compared to other movements, a more distinctive pattern may have been obtained in the signals generated as a result of the movement and therefore in the HHS images compared to other movements. For this reason, it is thought that the features extracted from the images of this movement can be

distinguished better than other movements. It has been observed that the proposed classification system provided different accuracy values for different movements. In this case, the situation of misidentification and classification has an effect. Due to the nature of the movement, the number, and type of muscles involved in each movement, or the similarity between the movements vary. This may be the reason for the variability in classification performance. The variability can be minimized by providing EMG from a relevant and appropriate number of muscles, better preprocessing, and using a large dataset. For example, movements with the same primary activation muscles can be considered similar [53]. According to these, it is thought that the mentioned conditions may have an effect on misclassification cases.

The RMSE and MAE values in Table 3.5 show that the SC-ANFIS model classification estimates are closer to the experimental value compared to the FCM-ANFIS model. In the R correlation coefficients, since it is known that values close to 1 indicate a direct relationship between the variables, it reveals that the SC has a better relationship with the input data than the FCM. Similarly, in the coefficient of determination R^2 , which tests the suitability of the obtained regression equation to the data, it is seen that the SC method is superior to FCM for the test data.

4.2.1 Comparison with State-of-the-Art Studies

In order to evaluate the results of the proposed gesture recognition system, a comparison with some FL-based studies carried out in recent years was made. In order to obtain a complete, comprehensive, and fair comparison between FL-based classifier methods, Table 4.2 was created including the number of individuals, muscle types, number of hand movements, number of features used, and other factors. However, SC and FCM-based methods applied in this study were compared among themselves and then with other literature studies. The proposed method has been shown to be superior according to the complexity level, the number of hand gesture classes, the number of muscles, the number of features, and the number of people in the dataset. It is seen that this study is more comprehensive than most of the data used in the studies in Table 4.2, in terms of the dataset structure. However, when comparing the types of features used, other studies have extracted features directly from the time domain of the signal [26,53,54,56] or from TFRs [53]. In this study, an F1-Score of 94.38% (SC) and an

accuracy percentage of 93.88% (SC) are presented with the classification method performed using the features extracted from the TF spectrum images obtained after the signal analysis method (HHT). In the study of Balbinot and Favieiro [54], although the number of subjects was the same as in this proposed study and 8 electrodes were used, it is thought that there may be insufficient information in the data representation, since only the RMS value was used as a feature, and the classification was based on a single parameter. Although the study of Khezri and Jahed [53] gave high accuracy, it was insufficient in terms of the number of people and the number of electrodes like other studies. Because, as it is known, the increase in the number of muscles measured signal increases in direct proportion to the accuracy value [123].

Table 4.2: Comparison table of the current FL studies in literature

Study	Year	Classifier	Gesture	Subject	Channel	Feature	ACC (%)
[53]	2011	Sugeno (SC-ANFIS)	6	4	2	4	92.00
[54]	2013	Sugeno (Neuro-Fuzzy)	7	30	8	1	86.00
[26]	2015	Sugeno (FCM-ANFIS)	4	4	2	6	88.90
[55]	2016	Sugeno (ANFIS)	4	1	2	4	96.85
[56]	2017	Mamdani (Fuzzy Logic)	2	1	2	3	93.12
[57]	2018	Sugeno (ANFIS)	5	1	8	13	72.00
[58]	2020	Sugeno (ANFIS)	7	-	8	1	98.09
[124]	2022	Genetic Fuzzy Classifier	3	36	8	1	67.00
Proposed Method	2022	Sugeno (SC-ANFIS and FCM-ANFIS)	7	30	4	6	SC-ANFIS: 93.88 FCM-ANFIS: 92.10

When the other methods in Table 4.2 were examined in terms of the number of people, it was seen that only one person's data were used in the studies [55–57]. This situation has made the methods and results open to discussion, even though the accuracy values of the mentioned studies are high. The performance values obtained in these studies may only be suitable for modeling for personal use. Otherwise, the created model may

be insufficient to test new data. In the study of Arozi et al. [58], which classified the same number of movements as this study using only one feature, although a high accuracy was obtained, information about the number of subjects was not presented. In addition, various FL approaches applied in studies cause different results. Unlike other studies, the study of Ulkir et al. [56] used Mamdani-type fuzzy logic system. For this reason, fuzzy rules were created manually by considering features and movements. Although the result of the mentioned study is remarkable, ANFIS is preferable to create the rules in its own networks instead of manually creating the rules as applied in other approaches. In that point, since Mamdani type classification is not suitable for multiple datasets, Sugeno-based ANFIS facilitates system creation. In the study of Palmer and Cohen [124], unlike other studies, a genetic fuzzy classifier was developed, and three movements were classified, and 67.00% test ACC was obtained. However, that study also used the Mamdani type inference system and the test accuracy obtained confirms the knowledge that the Mamdani type classification is not suitable for multiple datasets. In addition, the performance values obtained are quite low compared to other studies.

The two clustering methods were used separately in previous studies (study [53] (SC) and study [26] (FCM)). In this study, clustering methods were used with the same dataset and thus a comparative approach was presented for two different methods. Khezri and Jahed [53] obtained higher accuracy using SC-ANFIS than Kaiser et al., who used FCM-ANFIS [55]. In this proposed study, a similar situation was tested on the same dataset. The results confirm that the SC-based method is more successful than FCM. The main reason for this situation; may be due to the fact that the FCM is stuck at local minimum points and cannot handle outliers in the data [125]. In addition, the obtained results support the usability of TF images created by HHT in classification studies.

4.3 Evaluation of DL-based Classification Results

In this section, the classification results of DL-based approaches are evaluated in four subtitles.

4.3.1 Evaluation of IMF and WL Results

Movement-based muscle contraction naturally increases the frequency and amplitude of EMG signals. In the HHS approach, distinct combinations are predominantly selected from the first IMFs due to the fact that the first IMFs extracted by the EMD algorithm consist of higher frequencies, and the frequency content decreases as the IMF number increases. Further, first-several normally distributed IMFs, have been utilized for feature extraction and classification in many studies employing EMD [2,118]. In observations of this study, the IMF₁-based HHS image represents the energy density at higher frequencies and provides more resolution compared to the IMF₃-based HHT image when examined in terms of both instantaneous energy and instantaneous frequency. The important point here is how the instantaneous energies may better represent the gesture in the HHS images. The effect of the IMF combinations and segment length on the classification performance was examined. In the individual IMF combinations (IMF₁, IMF₂, and IMF₃), the accuracy decreased as the IMF number increased. This indicates that the discrimination ability of the model performs better at higher frequencies. When multiple IMF combinations are used, classification performance is better than for all individual combinations. Thus, it turns out that IMFs perform better in multiple combinations. In the limit range for online tasks (150-250 ms WL), the best performance was obtained using IMF₁₋₁₀₋₃ at 250 ms. It should be noted that no statistical significance was found between both. The classification results of the selection of the best IMF combination are in line with the findings of state-of-the-art studies mentioned in the previous section. It is concluded that IMFs obtained by EMD should be tested to ensure whether are normally distributed before analyses, and it is recommended that only normally-distributed IMFs may be used in the HHT to preserve significant features in the TF domain. Furthermore, IMF₁ individually outperformed other individual IMFs at all WLs and the statistical tests revealed this significance. However, all multiple combinations of

IMFs yielded better performance than IMF_1 in all cases. Therefore, the results have emphasized the usage of a combination of IMFs rather than individually. Nevertheless, IMF selection tasks require specific attention and need to be further investigated.

Another parameter is the selection of signal segment length, which may directly affect classification performance. Recent studies [49,67,126] suggested 150-250 ms as the limits for use in the real-time task. While the attraction is to reduce the response time for actions of actuators in the real-time task, it brings the challenges of extracting distinct information from sEMG with smaller WLs. It should be noted that TFA methods provide significant information from signals while the resolution in TF images may be associated with sample size. Therefore, choosing the optimum segment length becomes important. The selection of the appropriate WL is also related to the used feature extraction method. Therefore, the effect of WL size should be investigated by considering the feasibility of online tasks, the high volume of data required by DL architectures, and the feature extraction method. The overall implication obtained in this study is that it would be appropriate to select the larger possible WL using a sliding window due to the TFA methods requiring larger WLs to increase resolution as well as DL architectures requiring larger volume data to increase classification performance. The obtained findings encourage the implications. As the sEMG segment length variation experiment is considered, the worst ACC values were yielded at the smaller WL of 200 ms. Moreover, it can be easily noticed the visual differences between the smaller WLs (Figure 3.5-(a)) and larger WLs (Figure 3.5-(b)). As such, WL directly affects the instantaneous energy density in TFA methods. Hence, the overall accuracy tends to increase as the window size extends. As a result, 250 ms WL is selected for use in TFA methods considering the online tasks and classification performance.

4.3.2 Evaluation of Results of CNN Models

It is also investigated the CNNs' performance. Among the CNN models in Table 3.6, SqueezeNet was the only architecture with an F1-score value below 80%. Nonetheless, no statistical significance ($p= 0.735$) was found between the AlexNet and SqueezeNet. The only advantage of SqueezeNet is the lightweight model compared to others. A significant validation accuracy difference of about 13% was between SqueezeNet and

ResNet-50. For all other metrics, SqueezeNet also presented worse performance than the other architectures. Similar ACCs were yielded for AlexNet. Even though the AlexNet has many trainable parameters, the performance was not yielded as other networks like VGGs due to its depth (8). The ACCs of GoogLeNet and VGG-19 were close to each other and outperformed VGG-16, but not as well as residual network-based architectures. VGG-19 obtained only 0.29% higher accuracy than GoogLeNet. Their AUC values are equal to each other but far from the ideal values. Also, there is a big difference in cross-entropy loss values compared to residual networks. As seen in Table 3.6, the highest classification results were obtained with the ResNet-based architectures. It was observed that ResNet-50 achieved better accuracy with a difference of 1.52% compared to ResNet-34. ResNet-50 achieved lower values for cross-entropy loss and MSE, providing better performance. They achieved an F1-score of above 93% and provided an above 3.21% significance from GooLeNet and VGG-19. But there is no statistical significance ($p=0.378$) was found between each other.

Although ResNets use fewer parameters compared to other architectures such as VGGs and AlexNet, they may offer better classification accuracy because of containing shortcuts and more layers in their structure. The overall ACC values increased as the network complexity and the number of layers further increased. It is concluded that traditional CV methods such as the SKCV may be useful to compare the performance of the feature extraction and classification approaches. Also, the architecture selection may be conducted with supporting from experimental outcomes and considering the complexity of the features, the volume of the dataset, and the number of classes. The results revealed the significance of using task-specific architecture and calculating robust metrics in performance evaluation. Eventually, the achieved better performance by residual network-based structures among other architectures was in line with recent studies [47,68,104].

4.3.3 Gesture-based Evaluation

Evaluating the average results of the HHT-based approach, the pre-trained ResNet-50 model yielded over 98% accuracy in classifying Rest, Extension, Flexion, and Ulnar Deviation. However, the average ACC values were about 90% for radial deviation, punch, and open hand, respectively. Some movements are misclassified and the

maximum miss-classification rates of 9.42% and 7.73% were obtained for punch and open hand, respectively. The reason may be that similar muscles exhibit similar activation when the EMG signals are measured during some gestures or dominant activation of multiple channels may be observed simultaneously during the punch and open hand movements. Therefore, this may be observed in the HHS images and lead to similarity, hence miss-classification. In addition, in this study's data collection procedure consisting of five repetitions, the participants' muscle fatigue in the last repetitions while performing the movements may have been effective in the activity of the muscles. As a result, no miss-classification occurred greater than 10% in a single gesture class.

4.3.4 Comparison with State-of-the-Art Studies

In Table 4.3, recent state-of-the-art studies using DL are summarized considering their best classification performance to compare with this work in terms of the number of classes, channels, participants, WLS, used deep architectures, the signal processing method of the input data, CV strategy, and accuracy. The performance of the deep model may be directly affected by the number of EMG channels, participants, WLS, and classified movements. The proposed method aims to enhance the classification performance while not increasing the training cost. In the study of Shanmuganathan et al. [127], EMG signals were collected from 2-channel. This may result in insufficient EMG signal information obtained from the muscles responsible for contraction and relaxation events because many muscles may be responsible for only one hand gesture. Thus, the classification performance may be reduced. Similarly, in the study of Savithri et al. [128], using a lower number of participants and fewer channels may have affected the classification performance. In the study of Nahid et al. [47], LSTM-ResNet-18 combination and CWT were used which achieved higher performance than the HHT-based approach but tested with a little amount of data in terms of the number of channels (2) and subjects (8). In the study of Oh and Jo [49], 8-channel EMG signals from only 3 participants were considered. Similarly, the number of participants and EMG channels directly affected the classification performance.

Table 4.3: Comparison table with recent DL studies

Study	Year	Mov.	Ch.	Subj.	WL	Network	Feature	CV	ACC (%)
[126]	2018	6	8	7	150	Single-Layer CNN	Raw EMG	k-fold	96.70
[129]	2019	7	8	17	260	ConvNet	CWT	k-fold	98.31
[130]	2019	6	8	20	50	Five-Layer CNN	PSD	TTS	99.00
[50]	2019	17	12	40	300	CNN+LSTM	Spectrogram	N/A	80.93
[131]	2020	6	16	10	N/A	Deep BP-LSTM	TD Features	TTS	92.00
[47]	2020	10	2	8	250	ResNet-18+LSTM	CWT	TTS	99.83
[51]	2020	2	4	52	N/A	GoogLeNet	XWT	TTS	97.60
[132]	2020	7	8	17	260	EMGNet	CWT	TTS	98.61
[133]	2020	6	8	6	500	CRNN	SAWT	TTS	92.50
[72]	2020	17	12	40	100	CNN+LSTM	EMG Signal	TTS	98.01
[127]	2020	4	2	10	1000	R-CNN	WPT	TTS	96.48
[128]	2021	6	2	1	N/A	4-Layer CNN	EMG Image	N/A	91.66
[49]	2021	6	8	3	200	2-Layer CNN	SAWT	TTS	93.90
[46]	2022	6	8	36	50	Hybrid CNN	EMG Signal	TTS	93.48
[134]	2022	29	10	27	100	DCN	Signal Image	TTS	79.54
This Study	2022	7	4	30	250	ResNet-50	HHT	SKCV	93.75

***Mov.:** Movement, **Ch.:** Channel, **Subj.:** Subject, **WL:** Window Length, **CV:** Cross-Validation, **TTS:** Train-Test Split, **R-CNN:** Region-based CNN, **DCN:** Deformable convolutional network, **WPT:** Wavelet Packet Transform. WL values are given as ms.

As the number of channels increases, more information about the movement may be obtained, but it is important to keep the channel number at an optimum level. Not only the processing of extra input causes a computational load on the system but also may cause the system overfitting. This may have resulted in lower than the expected ACC despite the use of multiple channels in the study of Wang et al. [131]. In order to find the optimum number of channels, it is necessary to determine the muscles responsible for hand movements obtained from the forearm. The four muscles responsible for the forearm gestures considered in this study were determined by expert physicians. Additionally, in the study of Roy et al. [51], 4-channel EMG data (similar to the proposed study) of 52 participants (more than in this study) was used, and better accuracy values were obtained when compared with the proposed study. However, it should be noted that the mentioned study classifies fewer movements (4) in contrast to the proposed work. In the study of Wang et al. [134] among the studies in Table 4.3, was considered the highest number of movements (29 movements). Also, they analyze 27 participants' EMG data using 10-channel information. As a result, worse ACC yielded with this combination. In the study of Karnam et al. [46], a classification task was conducted with 8-channel data of 36 subjects who performed 6 movements. The obtained ACC was about 93%. Besides the input data generation processes, choosing dataset parameters by considering mentioned acquisition criteria may increase both the success and effectiveness of the deep model.

In addition, when the HHT-based method is compared to the conventional TFAs, several TFA-based studies [23] state that HHT provides better results and speed than CWT in recognition cases. On the other hand, HHT depends on an adaptive mode decomposition approach and preserves the IF of each mode, resulting in an improved TF resolution. In this study, recorded EMG signals are represented in the TF domain with a high localization by using the EMD and HHT to capture the time and frequency characteristics of sEMG segments simultaneously. TF representations contain more significant information than the time-domain IMFs of the signals. Therefore, it is concluded that HHS images resulting from HHT, provide essential information for hand gesture classification. As a result, the trained models avoid high variance and overfitting. Moreover, conducted comprehensive experiments, a robust CV strategy, and statistical tests supported this study's claims.

Since sEMG data is collected from the skin surface with very low impedance, the quality of the collected signals is important to provide accurate results in classification. Signal collection tools have various parameters, from the filter applied to the sampling frequency. For instance, although it is a practical tool, the Myo-Armband (Thalmic Labs Inc.) has a low sampling frequency (e.g., 100 Hz or 200 Hz) and this may cause some limitations when data processing. It should be noted that the valuable information obtained from EMG signals is at higher frequency bands (up to 500 Hz) for gestures [112]. Also, its fixed wearable structure may cause inconsistency in fitting each user's relevant muscles. Therefore, studies in [126,129,130,132,133] using this data collection procedure may not include the more relevant information provided by the higher sampling rate and sensitive location of the muscles, as in the present study. Another important parameter is the size of the window when segmenting the sEMG signal. The widespread acceptance for real-time tasks is WL must be kept below the time delay of 250 ms [47,67]. The selection of WL is required specific attention. In Table 4.3, while some studies [50,127,129,132,133] conducted the analyses using WL above 250 ms, some of them [51,128,131] did not provide this important information. Moreover, in the studies [46] and [130], only 50 ms WL was selected. Although providing little delay time is attractive, considering the sample size only a 50 ms window may not consist of the distinct information for gesture classification. Therefore, it is investigated the effect of the two WL variations on classification performance in this study. According to the findings, HHT may be the lead in real-time applications considering the WL requirement, performance, and computational time.

Additionally, ResNet-50 outperforms a successful performance in the classification of TF spectra with a huge dataset among the other DL architectures. ResNets are different from conventional sequential architectures and ensure the training of networks with greater depth. As the layer's number rises, CNN's training becomes harder, and the ACC attains saturation, then starts to fall. During training in ResNet, it utilizes the short-cuts linkages straight commission the entry not solely to following contiguous but further to forth latter one. It describes the extracting of input features learned by that layer. Hence, the evanescence challenge of the gradients is averted with the re-utilizing activation of the preceding layer. In addition, the used ResNet-50 shows significantly lower training loss in all of these CNNs. Also, it has the ability to be

generalized with the validation data. This demonstrates that it addresses the degradation issue, and operates to acquire accuracy gains from rising deepness. This explains why ResNets have better classification performance than other CNN architectures in this study. Another aspect, as stated in the study of Kukker et al. [118], CNN-based approaches present a successful approach compared to ML techniques in problems that contain too much data and require diversity in the data, as in the proposed method.

Finally, almost none of the studies summarized in Table 4.3 evaluate test data with a robust CV technique. Many of them adopted arbitrary test data with a non-robust train-test split (TTS) method [28]. This situation may result in the trained models with bias and high variance. Also, no CV strategies were reported in the studies [50,128]. In a preliminary study of this thesis, EMD was applied to the EMG signals, and the first three IMF time-amplitude waveform images were separately granted as input to the DL architecture [2]. In the mentioned study, it was investigated which IMF carries more meaningful information in the EMG signals for hand gesture recognition. In the first three IMFs, a classification success of more than 97 was obtained. This aspect supports this study's IMF selection criteria. Nevertheless, no cross-validation strategy was carried out where the presented results show only training results. Therefore, the trained model suffers from uniformity and bias. In another preliminary study of this thesis [3], the STFT was utilized to originate sEMG spectrograms for the training of a deep network. Similarly, no cross-validation strategy was conducted, the trained models were not robust and also suffer from bias and uniformity. These studies examine the capability of TF images and the EMD method with deep learning techniques in hand gesture classification. In addition, while the traditional deep learning approach was adopted in both of the above preliminary studies, the advantage of transfer learning was used in this study. Hence, the training time and computational costs were reduced.

Here, the aim of the DL-based classification part is to take specific notice of the HHT method used for representing 1D EMG signals as 2D TF images. These images represent gesture-induced energy distribution and may provide close to ideal representations of the signal on the joint TF plane via HHT. Combining these TF representations with CNN's success in 2D image classification may allow for

promising gesture classification performance. Although the accuracy rates obtained in some of the above-mentioned studies were higher, the proposed method is more robust in testing new incoming data. It is concluded that the HHT method offers a promising and advantageous approach to creating informative images for the application of deep learning methods for 1D biological signals.

5 Conclusion

In this thesis, the performance of the TFA method in different AI approaches for gesture classification is presented. The high-resolution sEMG signals measured from the 4 forearm muscles of 30 subjects were recorded for the classification of 7 hand gestures. sEMG signals were converted to HHS images by applying HHT. These images and obtained features from them were classified into AI models. This study presents an alternative feature extraction approach based on image features obtained from GLCM of HHS images, which have not been used in hand gesture recognition for EMG-based AI systems before. The extracted six GLCM-based features were obtained using 200 ms-long sEMG signals. These features are classified into ML and FL models. In this respect, it has enabled the classification problem to be handled with a different and original point of view than other studies.

In ML results, by comparing the classification accuracy of two different feature extraction approaches, it is demonstrated that the GLCM-based approach is able to preserve the pattern embedded in the TFR of the sEMG signal related to the hand gesture. After PCA, five HHS image-based features yielded an accuracy of 90.87% from the Cubic SVM model, and it showed better performance than the GLCM features of snapshot images of EMG and IMF. Considering the reduced computational load in the training and prediction of the ML models, it is concluded that GLCM-based feature extraction from HHS images has remarkable success in the ML-based classification of hand gestures for myoelectric-based devices.

The proposed FL-based models differ from FL-based studies with similar datasets in that they use more features for classification. Six GLCM features obtained from HHS images were subjected to two different clustering methods and the success of clustering methods on classification was also examined. It was observed that the SC-

ANFIS (93.88%) based system achieved higher accuracy for recognizing each of the seven selected hand gestures compared to FCM-ANFIS (92.10%). When the classification performance is compared among the movements, it has been determined that the fist and flexion movements give more successful results than the other movements. In the proposed method, F1-score values for SC and FCM for seven hand movements were obtained as 94.38% and 92.49%, respectively. As a result, with the proposed study, a more comprehensive classification approach and evaluation is presented compared to the FL methods in the literature. It has been seen that the proposed method can be used effectively in the classification of complex hand movements. It has been shown that neural-fuzzy-based methods can provide good classification capability and be an alternative approach for non-invasive sEMG-based biological systems.

Unlike ML and FL, HHS images are used directly in DL models. The useful combination of HHT and TL is performed in the classification of hand gestures for the first time in the scope of this thesis. The spatial fusions of spectral images of sEMG segments obtained by the HHT were used for the training of various pre-trained CNN architectures. The effects of various WLs and different CNNs on classification performance are examined. IMF selection for HHS is also comprehensively discussed. The experimental results are evaluated with the SKCV strategy. The results revealed that the proposed method yields outstanding classification capability while not increasing the computational cost. The best performance of the HHT was achieved with an average F1-Score of 93.98% and average accuracy of 93.75% with the ResNet-50-based TL model using IMF_{1-10-3} and 250 ms-long sEMG segments within the real-time analysis limits. It is emphasized that the value of using normally distributed IMFs, and the selection of the largest possible WL to increase resolution in TF images. It is demonstrated that the HHT-TL combination has the potential to be an alternative feature extraction approach for the classification of non-stationary and nonlinear biosignals.

In this study, the use of HHT as TFA and its obtained HHS images and GLCM features of HHS images for gesture classification has provided an alternative classification approach to applications related to stationary and nonlinear biological signals that are difficult to analyze such as EMG. Although each model has its own positive and

negative aspects, it was the first time to use data from the same time-frequency analysis-based approach in three different AI models. When the results were compared in general, it was seen that the highest values were obtained with FL, and DL obtained a result very close to it. Although ML achieves lower accuracy than the other two methods, considering all processes, it has been seen to be a more advantageous method in terms of complexity and time. Because, while both FL and DL methods offer high accuracy, they perform more complex operations and require more time and hardware. Considering all this, machine learning that offers relatively low accuracy can be preferred by keeping complexity at an optimum level in real time applications. For higher accuracy, approaches such as DL can be applied.

It is a study that provides important information about the successful performance of the same TFA-based data, which is the Hilbert-Huang Spectra obtained by the Hilbert-Huang Transform. Each hand gesture classification model performed in three different artificial intelligence approaches using TFA has proven to be a successful and noteworthy alternative approach.

References

- [1] Katona J. A Review of Human–Computer Interaction and Virtual Reality Research fields in Cognitive InfoCommunications. *Applied Sciences* 2021; 11(6).
- [2] Kisa DH, Ozdemir MA, Guren O, Akan A. EMG based Hand Gesture Classification using Empirical Mode Decomposition Time-Series and Deep Learning. *Medical Technologies Congress, TIPTEKNO 2020. IEEE; 2020.* doi:10.1109/TIPTEKNO50054.2020.9299282.
- [3] Ozdemir MA, Kisa DH, Guren O, Onan A, Akan A. EMG based Hand Gesture Recognition using Deep Learning. *Medical Technologies Congress, TIPTEKNO 2020. IEEE; 2020.*
- [4] Shafiei SB, Durrani M, Jing Z, et al. Surgical hand gesture recognition utilizing electroencephalogram as input to the machine learning and network neuroscience algorithms. *Sensors* 2021; 21(5):1–18. doi:10.3390/s21051733.
- [5] Chen F, Deng J, Pang Z, Nejad MB, Yang H, Yang G. Finger angle-based hand gesture recognition for smart infrastructure using wearable wrist-worn camera. *Applied Sciences* 2018; 8(3). doi:10.3390/app8030369.
- [6] Jaramillo-Yáñez A, Benalcázar ME, Mena-Maldonado E. Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review. *Sensors* 2020; 20(9). doi:10.3390/s20092467.
- [7] Guo L, Lu Z, Yao L. Human-Machine Interaction Sensing Technology Based on Hand Gesture Recognition: A Review. *IEEE Transactions on Human-Machine Systems* 2021; 51(4):300–309. doi:10.1109/THMS.2021.3086003.

- [8] Osborn LE, Iskarous MM, Thakor N V. Sensing and control for prosthetic hands in clinical and research applications. *Wearable Robotics: Systems and Applications*. Elsevier; 2019:445–468.
- [9] Kisa DH, Yildirim MC, Ozdil B, Ozdemir MA, Guren O, Akan A. Investigating the Effect of Signal Channels and Features in Various Domains on the EMG-based Hand Gesture Classification. *Medical Technologies Congress, TIPTEKNO 2022*. IEEE; 2022.
- [10] Fajardo JM, Gomez O, Prieto F. EMG hand gesture classification using handcrafted and deep features. *Biomedical Signal Processing and Control 2021*; 63:102210. doi:10.1016/J.BSPC.2020.102210.
- [11] Fang C, He B, Wang Y, Cao J, Gao S. EMG-centered multisensory based technologies for pattern recognition in rehabilitation: State of the art and challenges. *Biosensors 2020*; 10(8). doi:10.3390/BIOS10080085.
- [12] Qi J, Jiang G, Li G, Sun Y, Tao B. Intelligent Human-Computer Interaction Based on Surface EMG Gesture Recognition. *IEEE Access 2019*; 7:61378–61387. doi:10.1109/ACCESS.2019.2914728.
- [13] Kamavuako EN, Scheme EJ, Englehart KB. On the usability of intramuscular EMG for prosthetic control: A Fitts' Law approach. *Journal of Electromyography and Kinesiology 2014*; 24(5):770–777. doi:10.1016/J.JELEKIN.2014.06.009.
- [14] McManus L, De Vito G, Lowery MM. Analysis and Biophysics of Surface EMG for Physiotherapists and Kinesiologists: Toward a Common Language With Rehabilitation Engineers. *Frontiers in Neurology 2020*; 11. doi:10.3389/fneur.2020.576729.
- [15] Rampichini S, Vieira TM, Castiglioni P, Merati G. Complexity analysis of surface electromyography for assessing the myoelectric manifestation of muscle fatigue: A review. *Entropy 2020*; 22(5). doi:10.3390/E22050529.

- [16] Ozdemir MA, Kisa DH, Guren O, Akan A. Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures. *Data in Brief*; 2022; 41:107921. doi:10.1016/J.DIB.2022.107921.
- [17] Ahsan MR, Ibrahimy MI, Khalifa OO. Electromyography (EMG) signal based hand gesture recognition using artificial neural network (ANN). 4th International Conference on Mechatronics: Integrated Engineering for Industrial and Societal Development, ICOM 2011; 2011.
- [18] Zawawi TNST, Abdullah AR, Jin WT, Sudirman R, Saad NM. Electromyography Signal Analysis Using Time and Frequency Domain for Health Screening System Task. *International Journal of Human and Technology Interaction* 2018; 2(1):2600–8122.
- [19] Too J, Abdullah AR, Saad NM. Classification of Hand movements based on discrete wavelet transform and enhanced feature extraction. *International Journal of Advanced Computer Science and Applications* 2019; 10(6):83–89. doi:10.14569/ijacsa.2019.0100612.
- [20] Karlsson S, Yu J, Akay M. Time-frequency analysis of myoelectric signals during dynamic contractions: A comparative study. *IEEE Transactions on Biomedical Engineering* 2000; 47(2):228–238. doi:10.1109/10.821766.
- [21] Frey CW, Sajidman M, Kuntze HB. Neuro-fuzzy supervisory control system for industrial batch processes. *IEEE International Conference on Fuzzy Systems*, vol 1. IEEE; 2000:116–121.
- [22] Huang NE, Shen Z, Long SR, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis; 1998.
- [23] Purwanto D, Chen YT, Fang WH. First-Person Action Recognition with Temporal Pooling and Hilbert-Huang Transform. *IEEE Transactions on Multimedia* 2019; 21(12):3122–3135. doi:10.1109/TMM.2019.2919434.
- [24] Gao T, Sheng W, Zhou M, Fang B, Luo F, Li J. Method for Fault Diagnosis of Temperature-Related MEMS Inertial Sensors by Combining Hilbert-Huang

- Transform and Deep Learning. *Sensors* 2020; 20(19):5633. doi:10.3390/S20195633.
- [25] Sugiyama M, Kawanabe M. *Machine Learning in Non-Stationary Environments*. The MIT Press; 2012.
- [26] Jahani Fariman H, Ahmad SA, Hamiruce Marhaban M, Ali Jan Ghasab M, Chappell PH. Simple and Computationally Efficient Movement Classification Approach for EMG-controlled Prosthetic Hand: ANFIS vs. Artificial Neural Network. *Intelligent Automation & Soft Computing* 2015; 21(4):559–573. doi:10.1080/10798587.2015.1008735.
- [27] Xie HB, Guo T, Bai S, Dokos S. Hybrid soft computing systems for electromyographic signals analysis: A review. *BioMedical Engineering Online* 2014; 13(1):1–19. doi:10.1186/1475-925X-13-8/TABLES/4.
- [28] Ozdemir MA, Degirmenci M, Izci E, Akan A. EEG-based emotion recognition with deep convolutional neural networks. *Biomedizinische Technik* 2021; 66(1):43–57. doi:10.1515/bmt-2019-0306.
- [29] Phinyomark A, Scheme E. EMG pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing* 2018; 2(3):1–27. doi:10.3390/bdcc2030021.
- [30] Ozdemir MA, Ozdemir GD, Guren O. Classification of COVID-19 electrocardiograms by using hexaxial feature mapping and deep learning. *BMC Medical Informatics and Decision Making* 2021; 21(1):1–20. doi:10.1186/S12911-021-01521-X/FIGURES/10.
- [31] Valdivieso Caraguay ÁL, Vásquez JP, Barona López LI, Benalcázar ME. Recognition of Hand Gestures Based on EMG Signals with Deep and Double-Deep Q-Networks. *Sensors* 2023; 23(8). doi:10.3390/s23083905.
- [32] Tepe C, Demir MC. Real-Time Classification of EMG Myo Armband Data Using Support Vector Machine. *IRBM* 2022; 43(4):300–308. doi:10.1016/J.IRBM.2022.06.001.

- [33] Briouza S, Gritli H, Khraief N, Belghith S, Singh D. EMG Signal Classification for Human Hand Rehabilitation via Two Machine Learning Techniques: kNN and SVM. 5th International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2022. IEEE; 2022:412–417.
- [34] Nunez-Montoya B, Valarezo Anazco M, Saravia-Avila A, Loayza FR, Valarezo Anazco E, Teran E. Supervised Machine Learning Applied to Non-Invasive EMG Signal Classification for an Anthropomorphic Robotic Hand. 2022 IEEE ANDESCON: Technology and Innovation for Andean Industry, ANDESCON 2022. IEEE; 2022.
- [35] Karapinar Senturk Z, Bakay MS. Machine Learning Based Hand Gesture Recognition via EMG Data. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 2021; 10(2). doi:10.14201/adcaij2021102123136.
- [36] Gouda MA, Wang H, Feng N. Accurate sEMG Classification for Hand Gesture Recognition Using Pattern Recognition Techniques. 7th International Conference on Information Science and Control Engineering, ICISCE 2020. IEEE; 2020:169–173.
- [37] Algüner AE, Ergezer H. Window length insensitive real-time EMG hand gesture classification using entropy calculated from globally parsed histograms. Measurement and Control 2023. doi:10.1177/00202940231153205.
- [38] Rabin N, Kahlon M, Malayev S, Ratnovsky A. Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques. Expert Systems with Applications 2020; 149:113281. doi:10.1016/J.ESWA.2020.113281.
- [39] Devaraj A, Nair AK. Hand Gesture Signal Classification using Machine Learning. IEEE International Conference on Communication and Signal Processing, ICCSP 2020. IEEE; 2020:390–394.
- [40] Shi WT, Lyu ZJ, Tang ST, Chia TL, Yang CY. A bionic hand controlled by hand gesture recognition based on surface EMG signals: A preliminary study.

- Biocybernetics and Biomedical Engineering 2018; 38(1):126–135. doi:10.1016/J.BBE.2017.11.001.
- [41] Azhiri RB, Esmacili M, Jafarzadeh M, Nourani M. EMG Signal Classification Using Reflection Coefficients and Extreme Value Machine. BioCAS 2021 - IEEE Biomedical Circuits and Systems Conference, Proceedings. IEEE; 2021.
- [42] Toro-Ossaba A, Jaramillo-Tigueros J, Tejada JC, Peña A, López-González A, Castanho RA. LSTM Recurrent Neural Network for Hand Gesture Recognition Using EMG Signals. Applied Sciences 2022; 12(19). doi:10.3390/app12199700.
- [43] Kim ES, Shin JW, Kwon YS, Park BY. EMG-Based Dynamic Hand Gesture Recognition Using Edge AI for Human–Robot Interaction. Electronics 2023; 12(7). doi:10.3390/electronics12071541.
- [44] Gopal P, Gesta A, Mohebbi A. A Systematic Study on Electromyography-Based Hand Gesture Recognition for Assistive Robots Using Deep Learning and Machine Learning Models. Sensors 2022; 22(10). doi:10.3390/s22103650.
- [45] Gozzi N, Malandri L, Mercurio F, Pedrocchi A. XAI for myo-controlled prosthesis: Explaining EMG data for hand gesture classification. Knowledge-Based Systems 2022; 240:108053. doi:10.1016/J.KNOSYS.2021.108053.
- [46] Karnam NK, Dubey SR, Turlapaty AC, Gokaraju B. EMGHandNet: A hybrid CNN and Bi-LSTM architecture for hand activity classification using surface EMG signals. Biocybernetics and Biomedical Engineering 2022; 42(1):325–340. doi:10.1016/J.BBE.2022.02.005.
- [47] Nahid N, Rahman A, Ahad MAR. Deep Learning Based Surface EMG Hand Gesture Classification for Low-Cost Myoelectric Prosthetic Hand. Joint 9th International Conference on Informatics, Electronics and Vision and 4th International Conference on Imaging, Vision and Pattern Recognition, ICIEV and icIVPR 2020. IEEE; 2020.

- [48] Güneş H, Akkaya AE. Using Wavelet Analysis and Deep Learning for EMG-Based Hand Movement Signal Classification. *Sakarya University Journal of Science* 1997; 27(1):214–225. doi:10.16984/saufenbilder.1176459.
- [49] Oh DC, Jo YU. Classification of Hand Gestures Based on Multi-channel EMG by Scale Average Wavelet Transform and Convolutional Neural Network. *International Journal of Control, Automation and Systems* 2021; 19(3):1443–1450. doi:10.1007/S12555-019-0802-1/METRICS.
- [50] Huang D, Chen B. Surface EMG Decoding for Hand Gestures Based on Spectrogram and CNN-LSTM. *2nd China Symposium on Cognitive Computing and Hybrid Intelligence, CCHI 2019*, 2019:123–126. doi:10.1109/CCHI.2019.8901936.
- [51] Roy SS, Samanta K, Chatterjee S, et al. Hand Movement Recognition Using Cross Spectrum Image Analysis of EMG Signals-A Deep Learning Approach. *National Conference on Emerging Trends on Sustainable Technology and Engineering Applications, NCETSTE A 2020*. IEEE; 2020.
- [52] Montazerin M, Zabihi S, Rahimian E, Mohammadi A, Naderkhani F. ViT-HGR: Vision Transformer-based Hand Gesture Recognition from High Density Surface EMG Signals. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, IEEE*; 2022:5115–5119.
- [53] Khezri M, Jahed M. A neuro-fuzzy inference system for sEMG-based identification of hand motion commands. *IEEE Transactions on Industrial Electronics* 2011; 58(5):1952–1960. doi:10.1109/TIE.2010.2053334.
- [54] Balbinot A, Favieiro G. A neuro-fuzzy system for characterization of arm movements. *Sensors* 2013; 13(2):2613–2630. doi:10.3390/s130202613.
- [55] Kaiser MS, Chowdhury ZI, Mamun S Al, Hussain A, Mahmud M. A Neuro-Fuzzy Control System Based on Feature Extraction of Surface Electromyogram Signal for Solar-Powered Wheelchair. *Cognitive Computation* 2016; 8(5):946–954. doi:10.1007/S12559-016-9398-4.

- [56] Ulkir O, Gokmen G, Kaplanoglu E. Emg Signal Classification Using Fuzzy Logic. *Balkan Journal of Electrical and Computer Engineering* 2017; 5(2):97–101. doi:10.17694/BAJECE.337941.
- [57] Caesarendra W, Tjahjowidodo T, Nico Y, Wahyudati S, Nurhasanah L. EMG finger movement classification based on ANFIS. *Journal of Physics: Conference Series* 2018; 1007(1). doi:10.1088/1742-6596/1007/1/012005.
- [58] Arozi M, Ariyanto M, Kristianto A, Munadi, Setiawan JD. EMG Signal Processing of Myo Armband Sensor for Prosthetic Hand Input using RMS and ANFIS. *7th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2020. IEEE*; 2020:36–40.
- [59] Wahid MF, Tafreshi R, Al-Sowaidi M, Langari R. Subject-independent hand gesture recognition using normalization and machine learning algorithms. *Journal of Computational Science* 2018; 27:69–76. doi:10.1016/J.JOCS.2018.04.019.
- [60] Küçük H, Eminoglu İ, Balcı K. Classification of neuromuscular diseases with artificial intelligence methods. *Journal of the Faculty of Engineering and Architecture of Gazi University* 2019; 34(4):1725–1741. doi:10.17341/gazimmfd.571506.
- [61] Kisa DH, Özdemir MA, Güren O, Alaybeyoğlu A. A decision-making mechanism based on EMG signals and adaptive neural fuzzy inference system (ANFIS) for hand gesture prediction. *Journal of the Faculty of Engineering and Architecture of Gazi University* 2023; 38(3):1417–1430. doi:10.17341/GAZIMMFD.1025221.
- [62] Ozdemir MA, Kisa DH, Guren O, Akan A. Hand gesture classification using time–frequency images and transfer learning based on CNN. *Biomedical Signal Processing and Control* 2022; 77:103787. doi:10.1016/J.BSPC.2022.103787.
- [63] Zhang X, Li R, Li H, Lu Z, Hu Y, Alhassan AB. Novel approach for electromyography-controlled prostheses based on facial action. *Medical and*

Biological Engineering and Computing 2020; 58(11):2685–2698.
doi:10.1007/S11517-020-02236-3/FIGURES/9.

- [64] Ahamed MA, Ahad MAU, Sohag MHA, Ahmad M. Development of low cost wireless biosignal acquisition system for ECG EMG and EOG. 2nd International Conference on Electrical Information and Communication Technologies, EICT 2015. IEEE; 2016:195–199.
- [65] Kisa DH, Ozdemir MA, Guren O, Akan A. Classification of Hand Gestures using sEMG Signals and Hilbert-Huang Transform. 30th European Signal Processing Conference, EUSIPCO 2022; 2022:1293–1297. doi:10.23919/EUSIPCO55093.2022.9909748.
- [66] Moin A, Zhou A, Rahimi A, et al. A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. Nature Electronics 2020 4:1 2020; 4(1):54–63. doi:10.1038/s41928-020-00510-8.
- [67] Lee KH, Min JY, Byun S. Electromyogram-Based Classification of Hand and Finger Gestures Using Artificial Neural Networks. Sensors 2022; 22(1). doi:10.3390/S22010225.
- [68] Ozdemir MA, Cura OK, Akan A. Epileptic EEG Classification by Using Time-Frequency Images for Deep Learning. International journal of neural systems 2021; 31(8):2150026. doi:10.1142/s012906572150026x.
- [69] Król-Józaga B. Atrial fibrillation detection using convolutional neural networks on 2-dimensional representation of ECG signal. Biomedical Signal Processing and Control 2022; 74. doi:10.1016/J.BSPC.2021.103470.
- [70] Altamirano-Altamirano A, Vera A, Leija L, Wolf D. Myoelectric signal analysis using Hilbert-Huang Transform to identify muscle activation features. 13th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2016. IEEE; 2016.
- [71] Andrade AO, Nasuto S, Kyberd P, Sweeney-Reed CM, Van Kanijn FR. EMG signal filtering based on Empirical Mode Decomposition. Biomedical Signal Processing and Control 2006; 1(1):44–55. doi:10.1016/J.BSPC.2006.03.003.

- [72] Rahimian E, Zabihi S, Atashzar SF, Asif A, Mohammadi A. Surface EMG-Based Hand Gesture Recognition via Hybrid and Dilated Deep Neural Network Architectures for Neurobotic Prostheses. *Journal of Medical Robotics Research* 2020; 05(01n02):2041001. doi:10.1142/S2424905X20410019.
- [73] Zahra A, Kanwal N, ur Rehman N, Ehsan S, McDonald-Maier KD. Seizure detection from EEG signals using Multivariate Empirical Mode Decomposition. *Computers in Biology and Medicine* 2017; 88:132–141. doi:10.1016/J.COMPBIOMED.2017.07.010.
- [74] Izci E, Özdemir MA, Sadighzadeh R, Akan A. Arrhythmia Detection on ECG Signals by Using Empirical Mode Decomposition. *Medical Technologies National Congress, TIPTEKNO 2018. IEEE*; 2018.
- [75] Andrade AO, Kyberd P, Nasuto SJ. The application of the Hilbert spectrum to the analysis of electromyographic signals. *Information Sciences* 2008; 178(9):2176–2193. doi:10.1016/J.INS.2007.12.013.
- [76] Hafizah WM, Supriyanto E, Yunus J. Feature extraction of kidney ultrasound images based on intensity histogram and gray level co-occurrence matrix. *6th Asia International Conference on Mathematical Modelling and Computer Simulation, AMS 2012*; 2012:115–120.
- [77] Haralick RM, Dinstein I, Shanmugam K. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics* 1973; SMC-3(6):610–621. doi:10.1109/TSMC.1973.4309314.
- [78] Haryanto T, Pratama A, Suhartanto H, Murni A, Kusmardi K, Pidanic J. Multipatch-GLCM for texture feature extraction on classification of the colon histopathology images using deep neural network with GPU acceleration. *Journal of Computer Science* 2020; 16(3):280–294. doi:10.3844/JCSSP.2020.280.294.
- [79] Priyanka, Kumar D. Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA. *Procedia Computer Science*, vol 167. Elsevier B.V.; 2020:1722–1731.

- [80] Vasantha M, Subbiah Bharathi D V, Dhamodharan R. Medical Image Feature, Extraction, Selection and Classification. *International Journal of Engineering Science and Technology* 2010; 2(6):2071–2076.
- [81] Cristianini Nello, Shawe-Taylor John. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press; 2000.
- [82] Chen C, Chen P, Belkacem AN, et al. Neural activities classification of left and right finger gestures during motor execution and motor imagery. *Brain-Computer Interfaces* 2020:1–11. doi:10.1080/2326263X.2020.1782124.
- [83] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. *Machine Learning* 1997; 29(2–3):131–163. doi:10.1023/A:1007465528199.
- [84] Singha J, Misra S, Laskar RH. Effect of variation in gesticulation pattern in dynamic hand gesture recognition system. *Neurocomputing* 2016; 208:269–280. doi:10.1016/J.NEUCOM.2016.05.049.
- [85] Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* 2015; 27(2):130. doi:10.11919/J.ISSN.1002-0829.215044.
- [86] Hand DJ (David J), Mannila Heikki, Smyth Padhraic. Principles of data mining. MIT Press; 2001.
- [87] Bhushan S, Alshehri M, Keshta I, Chakraverti AK, Rajpurohit J, Abugabah A. An Experimental Analysis of Various Machine Learning Algorithms for Hand Gesture Recognition. *Electronics* 2022; 11(6). doi:10.3390/electronics11060968.
- [88] Rosalina, Yusnita L, Hadisukmana N, Wahyu RB, Roestam R, Wahyu Y. Implementation of real-time static hand gesture recognition using artificial neural network. 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017. IEEE; 2018:1–6.

- [89] Tsai CY, Lee YH. The parameters effect on performance in ANN for hand gesture recognition system. *Expert Systems with Applications* 2011; 38(7):7980–7983. doi:10.1016/J.ESWA.2010.12.086.
- [90] Karlik B, Tokhi MO, Alci M. A Fuzzy Clustering Neural Network Architecture for Multifunction Upper-Limb Prosthesis. *IEEE Transactions on Biomedical Engineering* 2003; 50(11):1255–1261. doi:10.1109/TBME.2003.818469.
- [91] Zadeh LA. Fuzzy sets. *Information and Control* 1965; 8(3):338–353. doi:10.1016/S0019-9958(65)90241-X.
- [92] Jang JSR. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man and Cybernetics* 1993; 23(3):665–685. doi:10.1109/21.256541.
- [93] Al-Mayyahi A, Wang W, Birch P. Adaptive neuro-fuzzy technique for autonomous ground vehicle navigation. *Robotics* 2014; 3(4):349–370. doi:10.3390/ROBOTICS3040349.
- [94] Jia G, Lam HK, Ma S, Yang Z, Xu Y, Xiao B. Classification of Electromyographic Hand Gesture Signals Using Modified Fuzzy C-Means Clustering and Two-Step Machine Learning Approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2020; 28(6):1428–1435. doi:10.1109/TNSRE.2020.2986884.
- [95] Abdulshahed AM, Longstaff AP, Fletcher S, Myers A. Thermal error modelling of machine tools based on ANFIS with fuzzy c-means clustering using a thermal imaging camera. *Applied Mathematical Modelling* 2015; 39(7):1837–1852. doi:10.1016/J.APM.2014.10.016.
- [96] Yang MS, Nataliani Y. Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. *Pattern Recognition* 2017; 71:45–59. doi:10.1016/J.PATCOG.2017.05.017.
- [97] Hoerer T, Kuenzer C. Object detection and image segmentation with deep learning on Earth observation data: A review-part I: Evolution and recent trends. *Remote Sensing* 2020; 12(10).

- [98] Lindsay GW. Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *Journal of Cognitive Neuroscience* 2021; 33(10):2017–2031. doi:10.1162/JOCN_A_01544.
- [99] Theckedath D, Sedamkar RR. Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. *SN Computer Science* 2020; 1(2):1–7. doi:10.1007/S42979-020-0114-9/METRICS.
- [100] Khan A, Sohail A, Zahoor U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review* 2020; 53(8):5455–5516. doi:10.1007/s10462-020-09825-6.
- [101] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 2017; 60(6):84–90. doi:10.1145/3065386.
- [102] Jadhav P, Rajguru G, Datta D, Mukhopadhyay S. Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network. *Biocybernetics and Biomedical Engineering* 2020; 40(1):494–504. doi:10.1016/J.BBE.2020.01.010.
- [103] Mukhlif AA, Al-Khateeb B, Mohammed MA. Incorporating a Novel Dual Transfer Learning Approach for Medical Images. *Sensors* 2023; 23(2). doi:10.3390/s23020570.
- [104] Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine* 2020; 121:103795. doi:10.1016/J.COMPBIOMED.2020.103795.
- [105] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR; 2014.

- [106] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. 5th International Conference on Learning Representations; 2017.
- [107] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015.
- [108] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.
- [109] Kim YH, Park JB, Chang MS, Ryu JJ, Lim WH, Jung SK. Influence of the depth of the convolutional neural networks on an artificial intelligence model for diagnosis of orthognathic surgery. *Journal of Personalized Medicine* 2021; 11(5). doi:10.3390/JPM11050356.
- [110] Rao AS, Nguyen T, Palaniswami M, Ngo T. Vision-based automated crack detection using convolutional neural networks for condition assessment of infrastructure. *Structural Health Monitoring* 2021; 20(4):2124–2142. doi:10.1177/1475921720965445.
- [111] Pawara P, Okafor E, Surinta O, Schomaker L, Wiering M. Comparing local descriptors and bags of visualwords to deep convolutional neural networks for plant recognition. 6th International Conference on Pattern Recognition Applications and Methods, ICPRAM 2017. SciTePress; 2017:479–486.
- [112] Li G, Li Y, Yu L, Geng Y. Conditioning and sampling issues of EMG signals in motion recognition of multifunctional myoelectric prostheses. *Annals of Biomedical Engineering* 2011; 39(6):1779–1787. doi:10.1007/S10439-011-0265-X/METRICS.
- [113] Hua S, Sun Z. A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. *Journal of Molecular Biology* 2001; 308(2):397–407. doi:10.1006/JMBI.2001.4580.
- [114] Englehart K, Hudgins B, Parker PA, Stevenson M. Classification of the myoelectric signal using time-frequency based representations. *Medical*

- Engineering & Physics 1999; 21(6–7):431–438. doi:10.1016/S1350-4533(99)00066-1.
- [115] Li K, Hogrel JY, Duchêne J, Hewson DJ. Analysis of fatigue and tremor during sustained maximal grip contractions using Hilbert-Huang Transformation. *Medical Engineering & Physics* 2012; 34(7):832–840. doi:10.1016/J.MEDENGGPHY.2011.09.025.
- [116] Huang H, Li T, Bruschini C, et al. EMG pattern recognition using decomposition techniques for constructing multiclass classifiers. *IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE Computer Society; 2016:1296–1301.
- [117] Sattar NY, Kausar Z, Usama SA, Farooq U, Khan US. EMG Based Control of Transhumeral Prosthesis Using Machine Learning Algorithms. *International Journal of Control, Automation and Systems* 2021; 19(10):3522–3532. doi:10.1007/S12555-019-1058-5/METRICS.
- [118] Kukker A, Sharma R, Malik H. Forearm movements classification of EMG signals using Hilbert Huang transform and artificial neural networks. *IEEE 7th Power India International Conference, PIICON 2016*. IEEE; 2016.
- [119] Benalcázar ME, Jaramillo AG, Zea JA, Paéz A, Andaluz VH. Hand gesture recognition using machine learning and the myo armband. *25th European Signal Processing Conference, EUSIPCO 2017*, IEEE; 2017:1040–1044.
- [120] T. Mahmood N, H. Al-Muifraje M, K. Salih S, R. Saeed T. Pattern Recognition of Composite Motions based on EMG Signal via Machine Learning. *Engineering and Technology Journal* 2021; 39(2A):295–305. doi:10.30684/etj.v39i2a.1743.
- [121] Qi J, Jiang G, Li G, Sun Y, Tao B. Surface EMG hand gesture recognition system based on PCA and GRNN. *Neural Computing and Applications* 2020; 32(10):6343–6351. doi:10.1007/S00521-019-04142-8/METRICS.
- [122] Laksono PW, Kitamura T, Muguro J, Matsushita K, Sasaki M, Amri Bin Suhaimi MS. Minimum mapping from EMG signals at human elbow and

- shoulder movements into two DoF upper-limb robot with machine learning. *Machines* 2021; 9(3):1–13. doi:10.3390/machines9030056.
- [123] Khushaba RN, Kodagoda S, Takruri M, Dissanayake G. Toward improved control of prosthetic fingers using surface electromyogram (EMG) signals. *Expert Systems with Applications* 2012; 39(12):10731–10738. doi:10.1016/J.ESWA.2012.02.192.
- [124] Palmer H, Cohen K. Genetic Fuzzy Hand Gesture Classifier. *Lecture Notes in Networks and Systems* 2022; 258:332–342. doi:10.1007/978-3-030-82099-2_30/COVER.
- [125] Fatih Eker A, Cil I. Capacitated vehicle routing problem with simulated annealing algorithm with initial solution improved with fuzzy c-means algorithm. *Journal of the Faculty of Engineering and Architecture of Gazi University* 2022; 37(2):783–798. doi:10.17341/gazimmfd.784653.
- [126] ur Rehman MZ, Waris A, Gilani SO, et al. Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques. *Sensors* 2018; 18(8). doi:10.3390/S18082497.
- [127] Shanmuganathan V, Yesudhas HR, Khan MS, Khari M, Gandomi AH. R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals. *Neural Computing and Applications* 2020; 32(21):16723–16736. doi:10.1007/S00521-020-05349-W/METRICS.
- [128] Savithri CN, Priya E, Sudharsanan J. Classification of sEMG Signal-Based Arm Action Using Convolutional Neural Network. *Signal and Image Processing Techniques for the Development of Intelligent Healthcare Systems* 2021:241–259. doi:10.1007/978-981-15-6141-2_13/COVER.
- [129] Côté-Allard U, Fall CL, Drouin A, et al. Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2019; 27(4):760–771. doi:10.1109/TNSRE.2019.2896269.

- [130] Pinzón-Arenas JO, Jiménez-Moreno R, Herrera-Benavides JE. Convolutional Neural Network for Hand Gesture Recognition using 8 different EMG Signals. 22nd Symposium on Image, Signal Processing and Artificial Vision, STSIVA 2019. IEEE; 2019:1–5.
- [131] Wang Y, Wu Q, Dey N, Fong S, Ashour AS. Deep back propagation–long short-term memory network based upper-limb sEMG signal classification for automated rehabilitation. *Biocybernetics and Biomedical Engineering* 2020; 40(3):987–1001. doi:10.1016/J.BBE.2020.05.003.
- [132] Chen L, Fu J, Wu Y, Li H, Zheng B. Hand gesture recognition using compact CNN via surface electromyography signals. *Sensors* 2020; 20(3). doi:10.3390/s20030672.
- [133] Jo YU, Oh DC. Real-Time Hand Gesture Classification Using Crnn with Scale Average Wavelet Transform. *Journal of Mechanics in Medicine and Biology* 2020; 20(10). doi:10.1142/S021951942040028X.
- [134] Wang H, Zhang Y, Liu C, Liu H. sEMG based hand gesture recognition with deformable convolutional network. *International Journal of Machine Learning and Cybernetics* 2022; 13(6):1729–1738. doi:10.1007/S13042-021-01482-7/METRICS.

Appendix

Publications from the Thesis

Conference Papers

1. Kisa, D. H., Ozdemir, M. A., Guren, O., & Akan, A. (2022, August). Classification of Hand Gestures using sEMG Signals and Hilbert-Huang Transform. In 2022 30th European Signal Processing Conference (EUSIPCO) (pp. 1293-1297). IEEE.

Journal Articles

1. Kisa, D. H., Ozdemir, M. A., Guren, O., & Alaybeyoglu, A. (2023). A decision-making mechanism based on EMG signals and adaptive neural fuzzy inference system (ANFIS) for hand gesture prediction. In *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38(3), 1417-1430.

2. Ozdemir, M. A., Kisa, D. H., Guren, O., & Akan, A. (2022). Hand gesture classification using time–frequency images and transfer learning based on CNN. *Biomedical Signal Processing and Control*, 77, 103787.

Projects

1. İzmir Kâtip Çelebi University Scientific Research Projects Coordination Unit - GAP Project (under grant 2022-GAP-MÜMF-0001)
2. The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2210-C National MSc/MA Scholarship Program in the Priority Fields in Science and Technology (under grant 1649B022103607)

Curriculum Vitae

Name Surname : Deniz Hande Kİsa

Education :

2015–2020 İzmir Kâtip Çelebi University, Dept. of Biomedical Eng. (BSc)

2020–2023 İzmir Kâtip Çelebi University, Dept. of Biomedical Eng. (MSc)

Work Experience:

2022 – 2023 Project Assistant in GAP Project (2022-GAP-MÜMF-0001)
İzmir Kâtip Çelebi University Scientific Research Projects
Coordination Unit

2021 – 2023 Researcher in Award Project (2021-ÖDL-MÜMF-0004)
İzmir Kâtip Çelebi University Scientific Research Projects
Coordination Unit

2021 – 2023 Scholarship student in TÜBİTAK 3501 Project (120E512)

2021 – 2022 Researcher in TÜBİTAK 2210-C National MSc/MA Scholarship
Program in the Priority Fields in Science and Technology
(1649B022103607)

Publications:

1. Kisa, D. H., Ozdemir, M. A., Guren, O., & Alaybeyoglu, A. (2023). A decision-making mechanism based on EMG signals and adaptive neural fuzzy inference system (ANFIS) for hand gesture prediction. In *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38(3), 1417-1430.
doi.org/10.17341/gazimmfd.1025221
2. Kisa, D. H., Yildirim, M. C., Ozdil, B., Ozdemir, M. A., Guren, O., & Akan, A. (2022, October). Investigating the Effect of Signal Channels and Features in Various Domains on the EMG-based Hand Gesture Classification. In *2022 Medical Technologies Congress (TIPTEKNO)* (pp. 1-5). IEEE.
doi.org/10.1109/TIPTEKNO56568.2022.9960235

3. Saripinar, L., Kisa, D. H., Ozdemir, M. A., & Guren, O. (2022, September). Statistical Analysis of Time-Frequency Features Based on Multivariate Synchrosqueezing Transform for Hand Gesture Classification. arXiv preprint arXiv:2209.13350.
4. Kisa, D. H., Ozdemir, M. A., Guren, O., & Akan, A. (2022, August). Classification of Hand Gestures using sEMG Signals and Hilbert-Huang Transform. In 2022 30th European Signal Processing Conference (EUSIPCO) (pp. 1293-1297). IEEE.
doi.org/10.23919/EUSIPCO55093.2022.9909748
5. Ozdemir, M. A., Kisa, D. H., Guren, O., & Akan, A. (2022). Hand gesture classification using time–frequency images and transfer learning based on CNN. *Biomedical Signal Processing and Control*, 77, 103787.
doi.org/10.1016/j.bspc.2022.103787
6. Ozdemir, M. A., Kisa, D. H., Guren, O., & Akan, A. (2022). Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures. *Data in brief*, 41, 107921.
doi.org/10.1016/j.dib.2022.107921
7. Kisa, D. H., Ozdemir, M. A., Guren, O., & Akan, A. (2020, November). EMG based hand gesture classification using empirical mode decomposition time-series and deep learning. In 2020 Medical Technologies Congress (TIPTEKNO) (pp. 1-4). IEEE.
doi.org/10.1109/TIPTEKNO50054.2020.9299282
8. Ozdemir, M. A., Kisa, D. H., Guren, O., Onan, A., & Akan, A. (2020, November). EMG based hand gesture recognition using deep learning. In 2020 Medical Technologies Congress (TIPTEKNO) (pp. 1-4). IEEE.
doi.org/10.1109/TIPTEKNO50054.2020.9299264