**ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL**

# HAND GESTURE RECOGNITION FOR TURKISH SIGN LANGUAGE USING ELECTROMYOGRAPHY FOR HUMAN-ROBOT INTERACTION

**M.Sc. THESIS**

**Mustafa SEDDIQI**

**Department of Computer Engineering**

**Computer Engineering Programme**

**JANUARY 2021**

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

# HAND GESTURE RECOGNITION FOR TURKISH SIGN LANGUAGE USING ELECTROMYOGRAPHY FOR HUMAN-ROBOT INTERACTION

**M.Sc. THESIS**

**Mustafa SEDDIQI**
**(504181526)**

**Department of Computer Engineering**

**Computer Engineering Programme**

**Thesis Advisor: Assoc. Prof. Dr. Hatice KÖSE**

**JANUARY 2021**

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

# İNSAN-ROBOT ETKİLEŞİMİ İÇİN ELEKTROMYOGRAFİ KULLANARAK TÜRK İŞARET DİLİ İÇİN EL HAREKETİ TANIMA

**YÜKSEK LİSANS TEZİ**

**Mustafa SEDDIQI**
**(504181526)**

**İbilgisayar Mühendisliği Anabilim Dalı**

**Computer Mühendisliği Programı**

**Tez Danışmanı: Assoc. Prof. Dr. Hatice KÖSE**

**OCAK 2021**

Mustafa SEDDIQI, a M.Sc. student of ITU Graduate School Engineering and Technology 504181526 successfully defended the thesis entitled "HAND GESTURE RECOGNITION FOR TURKISH SIGN LANGUAGE USING ELECTROMYOGRAPHY FOR HUMAN-ROBOT INTERACTION", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

| | | |
|---|---|---|
| **Thesis Advisor :** | **Assoc. Prof. Dr. Hatice KÖSE**<br>Istanbul Technical University | ..... |
| **Jury Members :** | **Prof. Dr. Şule Gündüz ÖĞÜDÜCÜ**<br>Istanbul Technical University | ..... |
| | **Assoc. Prof. Dr. Mehmet Fatih AMASYALI**<br>Yıldız Technical University | ..... |
| | | ..... |

**Date of Submission :** **22 January 2021**
**Date of Defense :** **22 February 2021**

***To my family,***

## FOREWORD

I believe the application of artificial intelligence and robots in rehabilitation and daily chores can improve life quality. This belief is the basis for my thesis and my Master's at the social robotics laboratory under Assoc. Prof. Hatice Köse's supervision.

After starting my Master's studies in September of 2018, I faced financial challenges in the following year. Thanks to my family for supporting my expenses to continue this journey and focus on my studies.

I would like to thank Dr. Hasan Kivrak for sharing his insightful ideas throughout this research.

The Deepest thanks go to my supervisor, Assoc. Prof. Hatice Köse, who helped me a lot in this journey by showing me how to approach research challenges. In light of her guidance, I was able to complete this research.

January 2021

Mustafa SEDDIQI
(Electrical and Computer Engineer)

**TABLE OF CONTENTS**

**Page**

## ABBREVIATIONS

**SL** : Sign Language
**TSL** : Turkish Sign Language
**ASL** : American Sign Language
**EMG** : Electromyography
**sEMG** : Surface Electromyography
**IMU** : Inertial Measurement Unit
**DTW** : Dynamic Time Warping
**HHM** : Hidden Markov Model
**MES** : Myoelectric Signal
**SVM** : Support Vector Machine
**MLP** : Multilayer Perceptron
**DWT** : Discrete Wavelet Transform
**PCA** : Principal Component Analysis
**HRI** : Human-Robot Interaction
**ROS** : Robot Operating System
**MEMS** : Micro-electromechanical Systems
**API** : Application Programming Interface
**ANN** : Artificial Neural Network
**ReLU** : Rectified Linear Unit
**GUI** : Graphical User Interface
**FT** : Fourier Transform
**DFT** : Discrete Fourier transform
**STFT** : Short Time Fourier Transform
**MAV** : Mean Absolute Value
**RMS** : Root Mean Square
**SSC** : Slope Sign Change
**WL** : Waveform Length
**HP** : Hjorth Parameters
**RF** : Random Forest
**t-SNE** : t-distributed stochastic neighbor embedding
**URDF** : Unified Robot Description Format

## LIST OF TABLES

**Page**

## LIST OF FIGURES

**Page**

## HAND GESTURE RECOGNITION FOR TURKISH SIGN LANGUAGE USING ELECTROMYOGRAPHY FOR HUMAN-ROBOT INTERACTION


### SUMMARY

Humans are socially intelligent beings they have complex connections to their environment. From the time they are born, children learn social skills from their experiences with family and other society members. Social skills have a positive correlation with one's social behavior, personal behavior, and cognitive ability. If a child could not develop these skills at an early age, he/she will have a problem in establishing social relationships and understanding a language (reading, writing, and grammar).

Hearing and speaking are the primary tools for humans to establish a connection with their environment. Deaf children, on the other hand, due to lack of hearing, cannot connect with the people around them like hearing children—this lack of communication results in hindering social skills and delay their cognitive development. If they cannot overcome this problem, it may even result in a permanent deficiency. Thus, a hearing-impaired child requires more support from their family, school, and society. Human-Robot Interaction (HRI) can compensate deaf children for learning language and learning social skills. Some deaf children may find humans' social behavior very complex; the level of interaction with the HRI can be adjusted according to a child.

This work's base motivation is to develop a sign recognizer model in an HRI platform to ease social skills learning in deaf children. Due to its unique linguistic properties and some unique signs, sign languages are different; in this work, we consider Turkish sign language (TSL), which is practiced in various Turkey regions.

Myo armband, a commercially available device with electromyography (EMG) and Inertial Measurement Unit (IMU) sensors, is used to recognize static signs of TSL. This armband is ready to use and can be easily connected with personal computers through a Bluetooth dongle.

Electromyography is the evaluation of the electrical signal in the muscle. The electrical signal is generated when the muscle is contracted. The electrical signal for EMG is recorded through electrodes. There are different EMG electrodes, such as needle-like, commonly used in electrodiagnostic testing, and surface-like, often used in rehabilitation. The EMG signal from muscle becomes noisy due to many factors such as skin disposition, the velocity of blood flow, skin warmth, muscle construction, and measuring.

IMU is a whole navigation device for measuring directions in three axes. Gyroscopes, accelerometers, and magnetometers are included as the inertial sensors. The angular rates from gyroscopes measure attitude, accelerometers outputs are accelerations for position determination, and magnetometers output are for the absolute orientation calculation.

We obtained EMG and IMUs data of ten healthy participants. Half of them were females of age 21-26 (average:23.8), and the other five were males of age 27-37 (average: 29.6). The data is recorded for stationary signs of TSL through Myo armband. Participants used their dominant hand, where nine used their right hand except for one female who used her left hand. The gestures were performed without contracting the muscles with excess strength, and the hand was moved in any direction of wrist, elbow, and shoulder joints.

We segmented the data into samples using the window technique with a size of 40 and 10 for EMG and IMUs, respectively.

Sign gestures are more correlated with orientation than gyroscope and accelerometer; thus, we did not use any fusion algorithm on IMUs, which can perform better on gyroscope and accelerometer values. IMUs were not preprocessed because the preprocessing does not increase the accuracy considerably in classifications. We used quaternion values for orientation, representing the position in four values rather than 3-axis Eular coordinates because they are more stable than other orientations. EMG noises, a wide range of noise resulting from inherent noise and movement artifacts, can be filtered significantly with a low-pass filter (allowing only low-frequency signals). Before applying the filter, we rectified the electromyography signal and then transformed it to frequency-domain using fast Fourier transform. The high-frequency signals are cleared, and the signal is transformed back to time-domain by inverse fast Fourier transform.

Features are extracted from the electromyography signals from both their time and time-frequency domains. This feature extraction is applied to the output of each electrode separately in addition to their cumulative sum. The features extracted from time-domain signals are waveform length, Hjorth parameters, mean, root mean square, and slope sign change. The signal is transformed through the Daubechies 5 wavelet (a fast implementation of discrete wavelet transform) into four layers. Subsequently, the features are obtained from the ratios of different transformed signal components from different layers. Furthermore, only the mean value is computed for each unit of the inertial sensors.

Before training any classifier, we standardize the inputs to have an equal contribution from all features; therefore, we normalized the input data with zero mean and unit variance. In this thesis, the accuracy metric is based on a cross-validation technique such that each time the data from one subject was fully considered the test set. The Machine Learning based classifiers; Multilayer Perceptron (MLP), and Random Forest (RF) are trained to classify hand gestures. The non-trainable parameters (hyper-parameters) of MLP (such as the loss function, number of layers, and regularization strength) and RF (such as selection criterion in trees, number of trees, and depth of trees) are tunned with random selection over a wide range of possible combinations.

The accuracy of these classifiers for all static signs of TSL was very low. Pretty much across all the signs, the RF classifier performed better in recognizing signs than MLP, so we advanced with RF.

When a gesture is performed, the envelope of electromyography signals varies for the same gesture performed by a different user. Additionally, the variation also exists between signals of two envelopes of a gesture performed by the same user. Thus,

constructing a general model that can be used by new users based on electromyography signals is very complicated. Our general model for all the static signs of Turkish Sign Language had a low accuracy score, but it gave us the intuition that which signs are more capable of being recognized with Myo armband. Furthermore, to assess electromyography and inertial sensors' characteristics, we selected five reliable signs and proceeded with our research.

The highest classification accuracy, 54.2%, is achieved when all the extracted features from electromyography and inertial sensors were used. The machine learning models were better at recognizing the gestures performed by males than the female participants. To scrutinize the discrepancy in accuracy score based on gender, we plotted the gestures in the two-dimensional plane using t-distributed stochastic neighbor embedding. The female gestures were more disperse than male gestures in the plots; we inferred that this is because the female subject had performed the signs with less muscular strength than the male. Therefore, training a machine learning classifier with reasonable accuracy on the females is more complicated.

The RF model trained solely on the male subjects for the five selected TSL gestures classifies with 78% accuracy. The signs ASL-O and ASL-L have generally performed with more muscular strength than their similar signs O-Hand and Thumb, respectively; thus, the model is less accurate in detecting the O-Hand and Thumb signs.

In this work, we controlled the hand's behavior of the simulated semi-humanoid robot Pepper with the signs recognized through the Myo armband. This work is a part of the ongoing RoboRehab (Assistive Audiology Rehabilitation Robot) project, where we aim to develop an effectively aware robotic rehabilitation platform to assist deaf children. In RoboRehab, we want to use different technologies and sensors to recognize TSL. Thus considering our project's aim, we used the robot operating system (ROS) to control the Pepper robot's behavior. Furthermore, because of inconsistency between drivers of Pepper's versions and the time limit, we used Gazebo, a robotic simulator, instead of a physical robot.

The TSL signs are very similar to each other with a little change in the fingers' position, a new sign is performed. Additionally, EMG signals have inter-class and intra-class variations for a gesture. These two issues make the detections of TSL signs with EMG sensors challenging. It is possible to detect TSL signs with EMG, but the accuracy and number of gestures are directly related to the electrodes' quality and number. Moreover, the gyroscope and accelerometer IMUs can recognize dynamic TSL signs but not static signs. The orientation IMUs can be efficient at classifying static TSL signs, but the sensor needs to be more accurate than the Myo armband's magnetometer IMU sensor.

The obtained results from this work motivate us to develop our model and the HRI platform further. In future works, we will try to overcome the gender problem and evaluate physical robots' success rate in children's learning processes.

# İNSAN-ROBOT ETKİLEŞİMİ İÇİN ELEKTROMYOGRAFİ KULLANARAK TÜRK İŞARET DİLİ İÇİN EL HAREKETİ TANIMA

## ÖZET


İnsanlar, çevreleriyle karmaşık bağlantıları olan sosyal olarak zeki varlıklardır. Çocuklar doğdukları andan itibaren sosyal becerileri aile ve diğer toplum üyeleriyle yaşadıkları deneyimlerden öğrenirler. Sosyal becerilerin, kişinin sosyal davranışı, kişisel davranışı ve bilişsel yeteneği ile pozitif bir ilişkisi vardır. Çocuk bu becerileri erken yaşta geliştiremezse, sosyal ilişkiler kurma ve bir dili (okuma, yazma ve dilbilgisi) öğrenmede sorun yaşar.

İşitme ve konuşma, insanların çevreleriyle bağlantı kurmaları için birincil araçlardır. Sağır çocuklar ise işitme eksikliğinden dolayı çevrelerindeki insanlarla işiten çocuklar gibi bağlantı kuramazlar- bu iletişim eksikliği sosyal becerilerin engellenmesine ve bilişsel gelişimlerinin gecikmesine neden olur. Bu sorunun üstesinden gelemezlerse bu onlar için kalıcı bir eksiklik olarak sonuçlanabilir. Bu nedenle, işitme engelli bir çocuk ailesinden, okulundan ve toplumundan daha fazla desteğe ihtiyaç duyar. İnsan-Robot Etkileşimi (HRI), işitme engelli çocukların dil öğrenmeleri ve sosyal becerileri öğrenmeleri için yardımcı bir araç olabilir. Bazı sağır çocuklar, insanların sosyal davranışlarını çok karmaşık bulabilirler ve bu yüzden HRI ile etkileşim düzeyi çocuğa göre ayarlanabilir.

Bu çalışmanın temel motivasyonu, sağır çocuklarda sosyal becerilerin öğrenilmesini kolaylaştırmak için HRI bir platformda işaret tanıyıcı model geliştirmektir. Eşsiz dil özellikleri ve bazı benzersiz işaretleri nedeniyle işaret dilleri farklıdır; Bu çalışmada sistemimize Türkiye'nin çeşitli bölgelerinde uygulanan Türk İşaret Dili (TSL) için uygun olarak geliştiriyoruz. Bu sistimimizde TSL'nin statik işaretlerini tanımak için Elektromiyografi (EMG) ve eylemsizlik ölçüm birimleri (IMU'lar) sensörleri ile çalışan Myo armband'ı kullanıyoruz. Bu kol bandının kullanımı rahattır ve bir Bluetooth aracılığıyla kişisel bilgisayarlara kolayca bağlanabilir.

EMG, iskelet kasının elektrik sinyalini değerlendiren bir tekniktir. Elektrik sinyali, kas kasıldığında üretilir. EMG için elektrik sinyali elektrotlar aracılığıyla kaydedilir. Elektrodiagnostik testlerde yaygın olarak kullanılan iğne benzeri ve genellikle rehabilitasyonda kullanılan yüzey benzeri farklı EMG elektrotları vardır. Kastan gelen EMG sinyali, cilt oluşumu, kan akış hızı, cilt sıcaklığı, doku yapısı ve ölçüm gibi birçok faktör nedeniyle gürültülü hale gelir. IMU, bir navigasyon sistemidir. IMU'da İvmeölçerler, jiroskoplar ve manyetometreler gibi bir eylemsiz sensör koleksiyonu dahildir. Jiroskop çıktıları, tutum ölçümü için kullanılan açısal hızlardır, ivmeölçer çıktıları konum belirleme için ivmelerdir ve manyetometreler mutlak yönelim hesaplamasını çıkarır. Myo kol bandı aracılığıyla sabit TSL işaretleri için on sağlıklı kişiden (21-26 yaşlarında 5 kadın, ortalama: 23.8; 27-37 yaşlarında 5 erkek, ortalama: 29.6) EMG ve IMU verileri elde ettik. Bir kadın dışında tüm katılımcılar sağ elini kullanıyordu. Hareketler omuz, dirsek ve bilek eklemlerinde serbestlik derecesi

(DOF) fark etmeksizin gerçekleştirildi. Verileri, EMG ve IMU'lar için sırasıyla 40 ve 10'luk ölçümlerde pencere tekniğini (window technique) kullanarak örneklere ayırıyoruz.

İşaret hareketleri yönelim ile jiroskop ve ivmeölçere göre daha fazla ilişkilidir; bu nedenle, IMU'larda jiroskop ve ivmeölçer değerlerinde daha iyi performans gösterebilen herhangi bir füzyon algoritması (fusion algorithm) kullanmadık. IMU değerlerinin önceden işlenmesi, sınıflandırmayı önemli ölçüde iyileştirmez. Bu nedenle, IMU değerlerini önceden işlemedik ve yönelim için kuaterniyon (quaternion) değerleri kullandık.

İçsel gürültü ve hareket kusurlarından kaynaklanan geniş bir gürültü aralığı olan EMG sesleri, bir düşük geçiş (low pass) filtresi ile önemli ölçüde filtrelenebilir. Alçak geçiren filtreyi uygulamak için, hızlı Fourier dönüşümü (FFT) kullanarak zaman-alanlı EMG sinyalini frekans-alanına dönüştürdük ve yüksek frekanslı sinyalleri filtreledikten sonra, ters FFT'yi kullanarak sinyali tekrar zaman alanına dönüştürdük. Dokuz kanaldan (8 EMG sensörleri ve toplamı) zaman-alanı ve zaman-frekans-alanı özelliklerini ve IMU'larda bir zaman-alanı özelliğini (ortalama) çıkardık. Bir örnekteki filtrelenmiş EMG sinyallerinden beş zaman-alanı özelliği çıkardık; ortalama mutlak değer (MAV), eğim işareti değişikliği (SSC), dalga formu uzunluğu (WL), ortalama kare kare (RMS) ve Hjorth parametreleri (HP). Bir örnekteki ham EMG sinyalinden öznitelikleri çıkarmak için dört katmanlı Daubechies 5 dalgacık ile ayrı dalgacık dönüşümleri kullandık.

Herhangi bir sınıflandırıcıyı eğitmeden önce, girdileri tüm özelliklerden eşit katkı sağlayacak şekilde standartlaştırıyoruz (sıfır ortalama ve 1 standart sapmaya sahip olacak şekilde dönüştürülür). Bu çalışmada, doğruluk puanını bildirmek için çapraz-doğrulama bir-denek-dışında-bırakmak prosedürü kullandık. Çok Katmanlı Algılayıcı (MLP) ve Rastgele Orman (RF) sınıflandırıcıları, veri kümesi üzerinde eğitilir. MLP'nin eğitilemeyen parametreleri; hiper-parametreler (kayıp fonksiyonu, katman sayısı ve düzenleme gücü gibi) ve RF (ağaçlarda seçim kriteri, ağaç sayısı ve ağaçların derinliği gibi) geniş bir olası kombinasyon yelpazesi üzerinden ayarlanıp seçildi.

Bu sınıflandırıcıların TSL'nin tüm statik işaretleri için doğruluğu çok düşüktü. Genel olarak, RF sınıflandırıcı MLP'den daha iyi çalıştı, bu yüzden RF ile devam ettik.

EMG sinyaliyle bir kez eğitilen ve herkes için çalışan genel bir model karmaşıktır çünkü EMG sinyal zarfları kullanıcılar arasında ve aynı kullanıcının içinde farklılıklar gösterir. EMG ve IMU'ların özelliklerini daha fazla araştırmak için beş tane daha güvenilir işareti seçtik.

EMG ve IMU'ların zaman alanı ve zaman-frekans alanı özelliklerini kullanarak elde edilen en yüksek doğruluk oranı %54,2'dir. Sınıflandırıcıların genel doğruluğu, kadınlara göre erkeklerde daha iyi performans göstermektedir. Erkekler ve kadınlar arasındaki farkı araştırmak için t-dağıtılmış stokastik komşu gömme (t-SNE) kullandık. Kadın hareketlerinin erkeklerden daha yaygın olduğunu bulduk. Bu nedenle kadınlar için kabul edilebilir doğrulukta herhangi bir sınıflandırıcı uyduramadık. Erkek denekler için, sınıflandırıcı %78 doğrulukla sonuçlandı. Sınıflandırıcı, bu işaretler arasında (O Hand, ASL O) ve (Thumb, ASL L) büyük bir doğrulukla ayırt edemez. Bu yanlış sınıflandırma, bir kavisteki işaretlerin birbirine çok benzer olmasıdır; akabinde, ASL O ve ASL L genellikle O Hand ve Thumb'dan daha fazla kas gücüyle yapılır.

Bu çalışmada, simüle edilmiş yarı insansı robot Pepper'ın elinin davranışını Myo armband'dan tanınan işaretlerle kontrol ettik.Bu çalışma, işitme engelli çocuklara yardımcı olmak için etkili bir şekilde bilinçli bir robotik rehabilitasyon platformu geliştirmeyi hedeflediğimiz, devam eden RoboRehab (Yardımcı Odyoloji Rehabilitasyon Robotu) projesinin bir parçasıdır. RoboRehab'da, TSL'yi tanımak için farklı teknolojiler ve sensörler kullanmak istiyoruz. Bu nedenle projemizin amacını göz önünde bulundurarak, Pepper robotunun davranışını kontrol etmek için Robot İşletim Sistemini (ROS) kullandık. Ayrıca, Pepper versiyonlarının sürücüleri ile süre sınırı arasındaki tutarsızlık nedeniyle, fiziksel bir robot yerine robotik bir simülatör olan Gazebo kullandık.

TSL işaretleri birbirine çok benziyor ve parmakların pozisyonunda küçük bir değişiklik olunca, yeni bir işaret anlamına geliyor. Ek olarak, EMG sinyalleri bir hareket için sınıflar arası ve sınıf içi varyasyonlara sahiptir. Bu iki sorun, EMG sensörleri ile TSL işaretlerinin tespit edilmesini zorlaştırır. TSL işaretlerini sEMG ile tespit etmek mümkündür, ancak hareketlerin doğruluğu ve sayısı doğrudan elektrotların kalitesi ve sayısı ile ilgilidir. Dahası, jiroskop ve ivmeölçer IMU'ları dinamik TSL işaretlerini tanıyabilirler ancak statik işaretleri ayırtamıyorlar. Oryantasyon IMU'ları statik TSL işaretlerini sınıflandırmada etkili olabilir, ancak sensörün Myo armband'ın manyetometre IMU sensöründen daha doğru olması gerekir.

Gelecekteki çalışmalarda bu sistemi insanların, özellikle işitme engelli bireylerin robotlarla iletişimini ve etkileşimini arrtırmak ve fiziksel robotların çocukların öğrenme süreçlerindeki başarı oranını değerlendirmeye çalışacağız.

# 1. INTRODUCTION

Hand gestures are a nonverbal communication medium used between humans. Intelligent and straightforward human-computer interaction can be created by understanding computer-based hand gesture recognition.

Recognition of hand movements serves as a guide to resolving many obstacles and providing human life with ease. In a wide range of applications, computers' capacity to comprehend human behaviors and their interpretation can be used. Sign language (SL) interpretation is one specific area of interest.

Recognition of gestures and SL involves the entire process of monitoring and recognizing the gestures executed and translating them into expressions and phrases that are semantically meaningful [1].

Recognition of static gestures of Turkish Sign Language (TSL) and providing interaction between hearing-impaired children and robots as recommended in [2–5] is the purpose of this thesis. This study is a part of the ongoing RoboRehab (Assistive Audiology Rehabilitation Robot) project, where we aim to develop an effectively aware robotic rehabilitation platform to assist deaf children. This project is funded under Grant number 118E214 by the Scientific and Technical Research Council of Turkey (TUBITAK). In RoboRehab, we want to use different technologies and sensors to recognize TSL signs. Therefore, in this thesis, we assess the usage of Electromyography signals and Inertial Measurement Units in sign recognition using the Myo armband.

## 1.1 Motivation

The social development process starts as a human is born by being a family member and becoming a society member. In this process, a child learns and experiences interactions with others that will later affect the child's social and personal behavior [6]. Children with hearing impairment suffer from establishing social relationships,

causing a lag in their cognitive skills growth. Deaf children do not lack the learning skill, but the sources and tools are not sufficient. Thus, they need more support from their families and schools.

Primary language must be learned at an early age, known as the critical period, to ease reading, writing, and grammar learning in children, else it can result in a permanent deficiency [7]. Human-Robot Interaction (HRI) can compensate for learning language problems by interacting with children through sign language or spoken language. Humans are socially intelligent beings, so their interactions are multi-model with different people based on their society. This social behavior of humans may seem very complicated to the children having a deficiency in communication skills. The robots' social behavior can be adjusted to different levels so the children can feel comfortable and learn communication skills at their own pace.

This thesis's base motivation is developing a wearable interface for gesture recognition to be used in HRI studies to help deaf children learn communication skills. Myo armband is cheap and easy to be used by children. Thus, we decided to use the Myo armband as the communication medium between robots and humans.

## 1.2 Sign Language

Most sign languages only include the top section of the body from the waist [8]. Additionally, when it is in various positions in the sentence, the same sign may significantly change meaning [9].

Hand gestures are distributed in multiple groups, including conversational, manipulative, controller, communicative groups. Sign language belongs to the branch of communicative gestures [10].

Most researches on SL recognition have been done on American Sign Language (ASL). The other common sign languages among researchers are Indian and Arabic. [1].

### 1.2.1 Turkish Sign Language

Turkish Sign Language is standard in all parts of Turkey with its unique symbols, words, and structure. Through hand gestures and facial expressions, TSL offers

non-verbal communication. It is used in Turkey, despite a particular regional dialectal variation. People with hearing impairments from various regions of Turkey can converse using TSL with each other [11].

The origin of Turkish sign language goes back to the 15th century when people with hearing impairments were called "mute" (dilsiz). In Fatih Sultan Mehmet's reigns, it was taken to the Ottoman Palace to provide information security and entertain the Sultan. At first, it was used only in the palace, and later, from the 18th, the signers were hired by Bab-ı Ali, where the hidden sessions took place at Meclis-i Has [12].

TSL has become a communication mechanism that, at the beginning of the 17th century, could state complex concepts. It was emphasized in the memories of Bobovius that, in the 17th century, TSL became a prestigious secret language [13].

TSL has its own different grammar and symbols; such as K, M, and 8 from ASL and third finger from Taiwanese SL cannot be understood by Turkish signers. This means that TSL has its own sequence of gestures [14]. TSL is a sequence of 36 signs which can be performed by both hands [15]. All of the 36 signs (Fig: 1.1) are considered stationary in this research.

### 1.3 Problem Definition

Recognition of hand gestures is an essential and challenging task. It has various applications like robotic hands [16], rehabilitation [17], sign detection [18], and controlling machines [19]. Cameras [20], smart gloves [21], electromyography [22], and inertial [23] sensors are used for this purpose. However, these sensors have their shortcomings; vision is highly dependent on lighting and orientation [24], hand gestures cannot be performed easily with gloves [25], and finally, electromyography and cheap inertial sensors output noisy signals [26].

Regarding the constraints and costs of the sensors available, we chose to evaluate the usage of Electromyography (EMG) and Inertial Measurement Unit (IMU) sensors. The data is collected through the Myo armband device, which has eight EMG and three inertial sensors. Myo armband is wireless, light, easy to use, and cheap. It is designed for hobbyists, so it is expected to generate noisy signals. In this research, we are trying to find processing methods to make the signals less noisy and applicable


**Figure 1.1** : Static signs of TSL.

in recognizing TSL signs. Furthermore, we want to control the robot with recognized signs. Robot Operating System (ROS) will be used to connect the robot with a gesture recognition model, which provides a standard way of developing robotic applications, so rewriting the code for a different robot will not be required.

The contribution of this work will be the usage of Myo armband's EMG and IMUs to detect TSL sign gestures and controlling robot behavior through ROS.

### 1.4 Overview of Thesis

Chapter 2, reviews some of the studies that provide a background for this thesis. This chapter provides a literature review of hand gestures recognition with visual sensors and gloves. Then it describes the works about the recognition of TSL. The second part

of this chapter discusses researches that have used EMG and Myo armband for gesture recognition.

Chapter 3 explains the proposed system. This chapter introduces the Myo armband and discusses the EMG and IMU signals in detail. Then the specification of the semi-humanoid robot Pepper is described. Next, this chapter explains the concept of the robot operating system (ROS). Furthermore, it outlines how some of the machine learning models work.

Chapter 4 describes the data collection and preprocessing methods. Then it thorogughly explains the feature extraction techniques in time and time-frequency doamins.

Chapter 5 explains the results obtained from the proposed system. Chapter 6 explains the finding from the TSL sign recognition model. Then discuss the relations between gesture recognition, EMG, and IMUs. Finally, chapter 7 provides conclusions and future work.

# 2. LITERATURE REVIEW

In this chapter, we review some of the studies that provide a background for this thesis. This chapter provides a literature review of hand gestures recognition with visual sensors and gloves. Then it describes the works about the recognition of TSL. The second part of this chapter discusses researches that have used EMG and Myo armband for gesture recognition.

Research in recognition of SL is greatly affected by researches in hand gesture recognition, as it is a group of communicating gestures. Consequently, hand gestures and SL recognition must be studied together [1].

Since the 20th century, the classification of signs has been a famous research topic [27]. There is much research about how chosen TSL signs and other sign languages are interpreted.

## 2.1 Recognition of Sign Language

Sign languages have distinct linguistic properties. Various studies on the identification of SLs have been carried out. Through vision, gloves, electromyography, and inertial sensors, gesture-recognition models collect data.

Some early gesture identification efforts go back to 1993; the authors in [28] used a speech recognition technique, Dynamic Time Warping (DTW), to detect hand movements using a camera. The research on French Sign Language, prologue and epilogue movements in [29] attempted to reduce using coarticulation parameters. Linguistic modeling helps create phrases using space, body animation's naturalness, non-manual features, and coarticulation.

In 1998 the work [30] used Hidden Markov Models (HMM) to recognize ASL signs using cameras. The first camera was looking directly at the person from the stationary point and achieved 90% accuracy, and the second camera was positioned on the head of the person, which resulted in 98% accuracy.

Work [31] has used Leap Motion to recognize and track finger movements of the hand. They have extracted features for different gestures and then experimented in the Unity game engine. The Leap Motion Controller's effective detection region is limited (0.25-60 cm above the device with 150 degrees field of view). Hand tracking information becomes unreliable when hands are near the borders of the detection area.

To translate SL to text, an intelligent model is developed using inertial, touch, and bend sensors mounted on a polyester-nylon glove [21]. It can translate 61 words of the Ecuadorian SL. A robotic hand-arm system, YoBu, commonly used in robotic teleoperation systems, is used for gesture recognition in [32]. They have analyzed the extracted features from the collected dataset.

A multi-class classification strategy based on Fisher score space, generated by Fisher kernels, is used to classify signs [33]. They have also proposed a method of adaptation based on Fisher score space selection to avoid retraining. In work [34] 98.75% accuracy has been achieved for eight defined gestures. The recognition of signs from images and videos to speech and vice versa is experimented by [35].

### 2.1.1 TSL recognition

A subset of selected primitive handshapes of TSL is recognized in work [36]. The work is using a method focused on the Leap Motion device. The authors have used heuristic, random forest, and multilayer perceptron classifiers; their heuristic model has achieved 99% accuracy.

The work [37] has used leap motion for detecting ten letters of TSL. They have achieved 100% accuracy with the deep neural network using feature extraction methods.

The authors in [38] used a data glove composed of strain, gyroscope, and accelerometer sensors to identify TSL numbers. A pair of strain sensors are attached to the thumb, index, and middle fingers, and only one sensor on the ring and little fingers. For some digits, the proposed classifier was successful.

TSL recognizer for 15 words/expressions is trained [39], Microsoft Kinect camera and Nui are used to record data from three individuals. The joint angles and joint

positions are used as features in the classifier. They have reported a high accuracy in this research.

## 2.2 Electromyography for Recognition of Hand Gestures

There is much work in recognizing hand gestures using electromyography, muscle signals when contracted. Using EMG signals provides a high degree of freedom. There are no occlusion and light sensitivity problems. Furthermore, it is not restricted to a narrow point of view; as long as the EMG signals are transmitted to a computer that can compute the signals and do the recognition.

EMG is used for different purposes, such as electrodiagnostic testing and rehabilitation equipment. Different EMG sensors are available, such as needle-like entering into the body, and surface-like are attached to the skin. Each type of EMG sensor has many variations that generate signals with different noises; thus, studies have been conducted based on EMG sensors' variants.

The work in [40] has studied dexterous and natural upper extremity prosthesis regulation using the myoelectric signal (MES). The device defined uses pattern recognition to process four MES channels, distinguishing between multiple limb movement groups. The method does not involve the MES data segmentation, enabling the transmission of a perpetual stream of the model results to a robotic hand. The EMG signal is used to detect abnormality in a muscle's contraction; injury, nerve damage, and neurological disorder cause muscle cells to not function properly [41]. In work [42], signal processing techniques on the EMG signal is used to diagnose grait disorder.

For rehabilitation purposes, EMG signals are often used in prosthetic hand control. The authors in [43] have used EMG and IMU sensors to detect 40 gestures offline and six in online recognition sessions. Then they have mapped six gestures to robotic arm actions to control the prosthetic hand. Eight EMG sensors are attached at the forearm (just below the elbow), and four other EMG sensors are connected to specific locations on the bicep. They have collected data from 20 healthy and two amputee participants but controlled the robotic arm with 11 healthy and one amputee participant.

A combination of EMG, IMUs, and electroencephalography is used to control a prosthetic hand's grasping. The prosthetic hand with five fingers was controlled with six motors (each finger with one motor but thump with two motors). The force sensor, potentiometers, and hall effect sensors on the prosthetic hand transmit information about the grasp to the controller, and the controller converts it into sound signals and sends it to the vibrators attached to the amputee's hand [44].

The EMG sensors have no limitations of vision sensors (such as illumination, occlusion, and point of view) and gloves (such as finger movements, hard to be used by children). Thus EMG sensors are also common in sign language recognition. Authors have developed a real-time sign language recognizer for ASL in [45] with EMG and IMU (gyroscope and accelerometer) sensors. They have achieved an accuracy below 40% for the 40 most used words in ASL. In their work, EMG and IMU together performed better than just EMG sensors, and out of four EMG sensors, the one placed on the forearm (above the palm) is the most important. They have developed their EMG sensors and attached them to a micro-controller to transmit the data via Bluetooth. They have trained multiple machine learning models, and the Support Vector Machine (SVM) has resulted at the highest accuracy.

The work [46] has developed a DTW and HMM to recognize 223 characters (hand movement during one or more performed gestures) of Chinese sign language. They have used four EMG sensors to classify 13 hand gestures and accelerometers to detect the four hand's motion (upward, downward, right, and left). The classification between two successive gestures (movement epenthesis) is done by the effective signals' start and end boundaries. Additionally, they have used accelerometers to detect whether the palm is upward (the hand will be then stationary) or the palm is downward (the hand can move in four directions) to ease the movements' classification. Feature extraction on the segment of both EMG signals and accelerations is used. The model's output combines 13 gestures with palm upward or hand is moving in four directions (13x5). The dataset is collected from five subjects, and the accuracy of the overall model has been reported around 95%.

The authors in [47] have used EMG sensors to detect nine gestures to control a human-computer interface and they have used nonstandard EMG sensors to achieve

92% accuracy. The [48] work has used EMG sensors (Bio Radio 150) and has achieved 82% accuracy for classifying 26 ASL signs.

### 2.2.1 Using Myo Armband for Recognition of Hand Gestures

The researchers often make their EMG sensors or prepare before attaching the sensors to the muscle. In such cases, information about the EMG sensors' performance is not provided, but the research focuses on EMG signals and their applications. However, some known devices, such as the Myo armband, are commercially available with EMG sensors. Myo armband has EMG and IMU sensors; it is wearable and transmits data to the PC through Bluetooth. It has paved the ground for the researchers to evaluate their works with other researchers. Thus, numerous researches are done on EMG and IMUs for gesture recognition using Myo armband.

Gated recurrent unit (GRU) network, deep learning model, is trained with the EMG signal of the Myo armband [49]. The data is collected from 35 able-bodied subjects for six very different gestures. The raw EMG data without preprocessing and feature extraction were standardized between -1 and 1 to feed the GRU network. The windowing technique (window-length;188, sliding-length; 20) has been used to segment the data. The window length (188 captured EMG data) is recorded in slightly less than one second with Myo armband, and the authors have estimated this is the amount of time required for making a static gesture correctly. The GRU network is composed of three GRU layers and one fully connected layer. The achieved accuracy is 77% when the model is tested with a new subject.

The authors in work [22] have trained an Artificial Neural Network (ANN) model to detect six quite diverse hand gestures. The classifier is based on the EMG signal of the Myo armband. In work, the data is collected from 12 individuals, each performing the gestures (five times in two seconds) in five sessions. The signal is rectified and segmented to create the dataset for training. From each segment of the data, five time-domain features are extracted. The authors have assessed the correctness of the classifier with various lengths of the window in segmentation. The model optimum window length has classified a new set of data from the same subjects with 98% accuracy.

Five different gestures is used in work [50]. The EMG of the Myo armband is used in this work for classification. The authors have collected data from 10 persons, such that each gesture was performed for two seconds. For each gesture, the participant started from the hand's rest position, performing the gesture, and going back to the rest position in two seconds. A gesture is performed 30 times by each subject to create the training-dataset; for the test-dataset, each subject performed the gesture six times in a different session. This way of performing gestures resulted in envelopes of EMG signals for each gesture. They have used the windowing technique (window-length;500 points, sliding-length; 10 points) to create samples, but any window with less than 100 points from the signal's envelope was discarded. Before extracting five time-domain features from each sample, the signal is rectified and filtered. A machine learning model, an MLP, is trained with features extracted from the training-dataset, and the model's performance to distinguish between test-dataset gestures is about 98%.

A one-vs-all SVM classifier has been trained to detect static signs from Brazilian sign language (LIBRAS) in work [51]. A subset of 20 signs that do not require any movement is selected from the alphabets and numbers. For the classification, only the EMG signal of the Myo armband is considered. The authors have used the windowing technique to create samples and rectified the signal before extracting the mean for each sensor. The data is collected from one subject, and the model has been tested on the same subject. The accuracy for the classification of signs lies in a wide range, such as R and U signs are classified correctly by 91% and 5%, respectively. The overall accuracy for all 20 signs is around 40%. The usage of IMUs for static signs is suggested for future work in classifying static signs through the Myo armband.

The EMG signal from the Myo armband is used to detect 16 gestures [52]. The dataset is obtained from five male individuals for ten Taiwanese number gestures, two wrist gestures, and four kinds of single finger extension. The participants performed each gesture for about 1 second, and the time interval from rest to performing the gesture and back to rest was about two seconds. Each subject repeated a gesture 800 times, and each repetition was considered as one sample. From time and time-frequency domains, two and five features are extracted from the signals, respectively. The signal is transformed to the time-frequency domain by Discrete Wavelet Transform

(DWT) technique. They have trained a Multilayer Perceptron (MLP) with the extracted features, and the model correctly classified the signs around 87%.

Sign recognizer for TSL is designed in work [53]. The authors have developed a model to classify 10 TSL digit gestures from each other when performed by the right hand. Myo armband is used for collecting EMG signals from one subject. The participant has performed each gesture for five seconds in five sessions. Then by applying the windowing technique (window-length;40, sliding-length;20), the data is segmented. From each segment of the data without any preprocessing, six features are obtained from each EMG sensor's time-domain signal. To decrease the number of features, the authors have used the principal component analysis (PCA) dimensionality reduction technique. They have trained an SVM classifier with 15 inputs (after applying PCA) with a segmented data subset. The model is tested with the remaining data set, and accuracy of 87% is reported.

To the authors' best knowledge, this work is the first research to focus on the recognition of all stationary TSL signs using electromyography and IMUs. We also want to develop an HRI system where the robot will understand and respond according to the performed sign.

# 3. PROPOSED SYSTEM

In this chapter, we provide the proposed methods. This chapter introduces the Myo armband and discusses the EMG and IMU signals in detail. Then the specification of the semi-humanoid robot Pepper is described. Next, this chapter explains the concept of the Robot Operating System (ROS). Furthermore, it outlines how some of the machine learning models work. Finally, our approach to preprocess and extract features from the data is discussed thoroughly.

## 3.1 Motivation

Social skill can only be learned when there is an interaction between child-environment and collaboration with others [54]. HRI can be developed with a focus on children's development in socialization and creativity. Social interaction with robots can be beneficial for different types of disabled children. The robot's effect in HRI changes to time, and more prolonged interactions are efficient than shorter interactions.

To design HRI to assess disabled children to have prolonged interactions with the robot, we want it to understand and act according to the child's hand gesture. We are using the Myo armband technology to recognize the hand gesture, which is generating EMG and IMU signals.

We want to evaluate EMG and IMU sensors' performance through the Myo armband in detecting TSL sing gestures in this work. Moreover, we want to connect this device to our HRI system. For this purpose, we are trying to find the answers to the following research questions:

- Can EMG be used for detecting TSL sign gestures?
- Can IMUs be used for detecting TSL sign gestures?
- Are the signals from the Myo armband good enough to detect TSL sign gestures?
- How to integrate our sign recognizer model with robots?

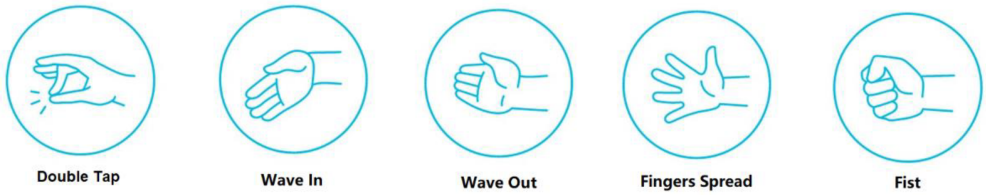



**Figure 3.1** : Default recognizable gestures of Myo armband [22]

**Figure 3.2** : Myo armband used in this thesis [56]

## 3.2 System Setup

### 3.2.1 Myo Armband

Technological advances in the manufacture of high-performance electronic devices frequently make them obtainable at more reasonable costs and develop their applicability in many applications today. Myo armband, which has provided a convenient interaction between human-computer, is used in this work. Myo armband is wireless, light, easy to use, and cheap (Fig: 3.2). It has eight dry EMG electrodes and three inertial sensors (9-axis IMU). The recording rate of Myo is 200 Hz for EMG signals and 50 Hz for IMU values [55]. Furthermore, it processes the signal with ARM Cortex-M4 microprocessor and transfers the sensed data through Bluetooth (Fig: 3.3) to other machines that interpret the signal and behave based on the interpretation [56].

The Myo armband performance compared to the common EMG sensors is studied in work [57]. The authors experimented with nine gestures on eight subjects. They have reported that mean recognition errors are 5.82 ± 3.63% and 9.86 ± 8.05% for regular EMG sensors and Myo armband, respectively.

The Myo comes installed with a set of recognizable hand gestures (Fig: 3.1) and these gestures can be used to move between slides in PowerPoint, navigate and scroll in web browsers through Myo-connect application.

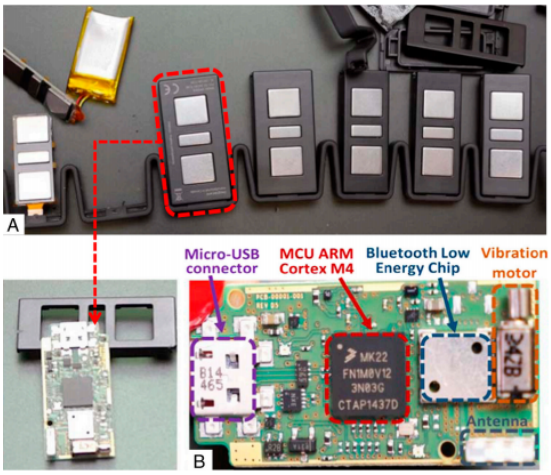



**Figure 3.3** : Myo armband (A) electrodes and (B) electronic control board [56]

### 3.2.2 Electromyography

Electromyography is the evaluation of the electrical signal generated in the muscle. Muscle signal is regulated by the nerve system and is generated by a group of motor units when the muscle is contracted [58]. The Motor unit is made of motor neurons with muscle fibers; a group of motor units usually contract a single muscle. Muscle contraction in physiology does not suggest shortening of muscle because, without changes in muscle length, muscle stress can be created, for example, holding something heavy at one location [59].

Compared to extracellular fluids around the muscle, the healthy muscle has a resting membrane potential of -80 mv. The membrane potential becomes more positive after injury or denervation, leading to an influx of Na+ into the weakened cell membrane [60].

MRI and x-ray are sophisticated photographs; however, EMG provides information in real-time about what is occurring physiologically to the nerve and the muscle [61]. The electrical signal for EMG is recorded through electrodes, requiring a ground, a reference, and an active electrode. There are needle and surface shape like electrodes.

**Needle Electrodes:** They are entered into the muscles and generally are disposable and are used on one person only. There are several kinds of needle electrodes, like Monopolar and Concentric needles. The Monopolar needle requires a surface electrode on the skin or a second needle in the subcutaneous tissue as reference. It is less expensive and less painful due to a narrow diameter. The Concentric needle is

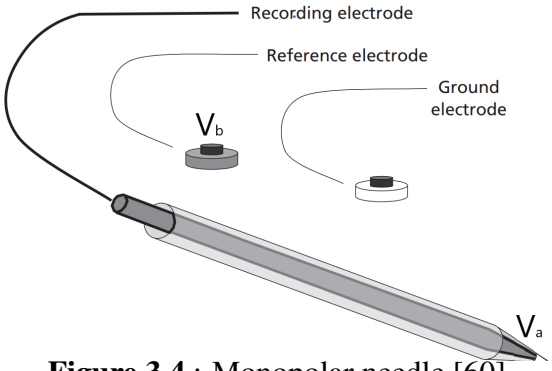


**Figure 3.4** : Monopolar needle [60]

referenced to its end, with no need for an additional electrode. It is electrically more stable than a Monopolar needle. Fig: 3.4 shows a Monopolar needle electrode with a surface electrode as a reference.

**Surface electrodes:** Usually, they have either ring or disk shape electrodes. They are either disposable or non-disposable as well. Without the need for tape or gel, disposable electrodes have a sticky underside that helps them adhere to the skin. The non-disposable electrodes are constructed from stainless steel, silver, or gold. They often need gel to minimize impedance and avoid artifacts due to abnormalities in the presence of skin and hair follicles. The non-disposal electrodes with no need for gel (dry electrodes) are recording much noise [49]. The recorded signal from surface electrodes can also be referred to as surface-EMG (sEMG). Fig: 3.5 illustrates measuring the electrical signal of muscle using surface electrodes used as active and reference. The benefit of these electrodes is easy usage that they can be used without hurting someone, so no need to anesthetize the user. The attaching is quick and painless. From this point forward, we are using EMG and sEMG interchangeably. The voltage difference in both Fig: 3.4 and Fig: 3.5 is calculated as in equation 3.1.

$$Vout = Va - Vb \tag{3.1}$$

The *Vout* signal passes first to the preamplifiers, then to the filters, then to the amplifier. Preamplifiers attenuate the biological signal to ensure that filters have better and more significant potential than system noise. Filters, both high pass and low pass try to block noise. Finally, amplifiers increase the magnitude of the signal through integrated circuits or chips to be displayed. At a healthy muscle, usually, the motor unit will fire in a regular pattern at about 5Hz.

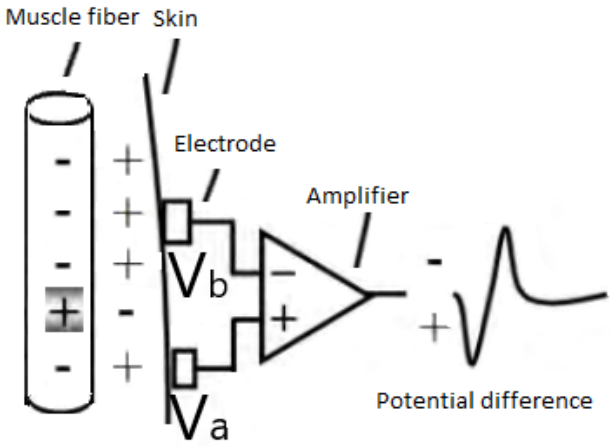



**Figure 3.5** : Electrical Signal measured using surface electrode [62]

EMG is used for electrodiagnostic testing, as rehabilitation equipment, and in generic Human-Computer Interface (HCI) and HRI applications. Electrodiagnostic testing is an essential method for physicians to distinguish between nerve and muscle disorders. Neuropathy is nerve damage that usually results in numbness, tingling, weakness, and pain in the affected region, while plexopath is a condition involving a network of nerves, blood vessels, or lymph vessels. Electrodiagnostic is used in the diagnosis of Carpal tunnel syndrome (neuropathy at the wrist), Ulnar neuropathy (neuropathy of the upper extremity), Radial neuropathy (humeral fracture or compression), Radiculopathy, Spinal stenosis, Peroneal neuropathy, Tarsal tunnel syndrome, Peripheral neuropathies, Myopathy (any disorder in the muscles), Brachial plexopathies, Lumbosacral plexopathies, and Motor neuron diseases [60].

The EMG signal from muscle becomes noisy due to many factors such as skin disposition, the velocity of blood flow, skin warmth, muscle construction, and measuring site [63]. The noises in EMG signals are as follows; electrode inherent noise [64] (every electronic device causes electrical noise and its frequency is between 0 and several thousand Hz), movement artifact [65] (due to the movement of the detection surface of electrode and cable; its frequency range is between 0 and 10 Hz), electromagnetic noise [66] (electromagnetic signals from the human body, the environment, or power sources; resulted noise from power sources can be 60 or 50 Hz frequency), cross talk [67] (due transference of signal from muscle to the electrode), internal noise [68] (due to anatomical, biochemical and physiological factors), the inherent instability of the signal [66] (due to motor firing rate which is quasi-random;

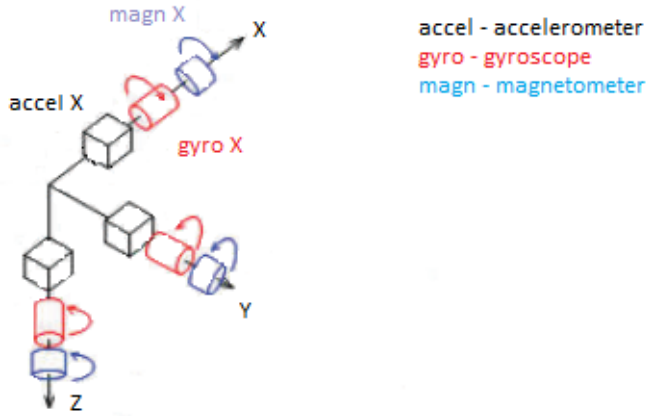



**Figure 3.6** : Body frame of inertial measurement unit [70]

its frequency range is between 0-20 Hz) and electrocardiographic artifacts [69] (due to the activity of the heart).

### 3.2.3 IMU

Inertial Measurement Unit (IMU) is a whole navigation device for measuring directions in three axes. Gyroscopes, accelerometers, and magnetometers are included as the inertial sensors. The angular rates from gyroscopes measure attitude, accelerometers outputs are accelerations for position determination, and magnetometers output are for the absolute orientation calculation [70].

**Gyroscope:** It is a 3-axis inertial sensor measuring angular rates. Different types of this sensor are used in strap-down IMUs like rotating-mass (based on conservation of angular momentum), optical (based on Sagnac effect), and vibratory (computes Corilios acceleration) [71].

**Accelerometer:** It is based on the second Newton's law of motion ($F = m.a$ where **F** is force, **a** is acceleration and **m** is mass which is constant) to measure acceleration in 3-axis. For strap-down IMUs, the pendulous (based on the deflection of hinged mass) and vibrating-beam (based on piezoelectric effect) type of accelerometers are used [72].

**Magnetometer:** It is used for absolute orientation calculation. It is based on Hall's impact theory. These sensors translate the magnetic field into the difference in output voltage [72].

Accelerometer and gyroscope inertial sensors are manufactured as micro-electromechanical systems (MEMS) sensors for lower cost IMUs. They are error-prone, but they are light and compatible in different applications [73].

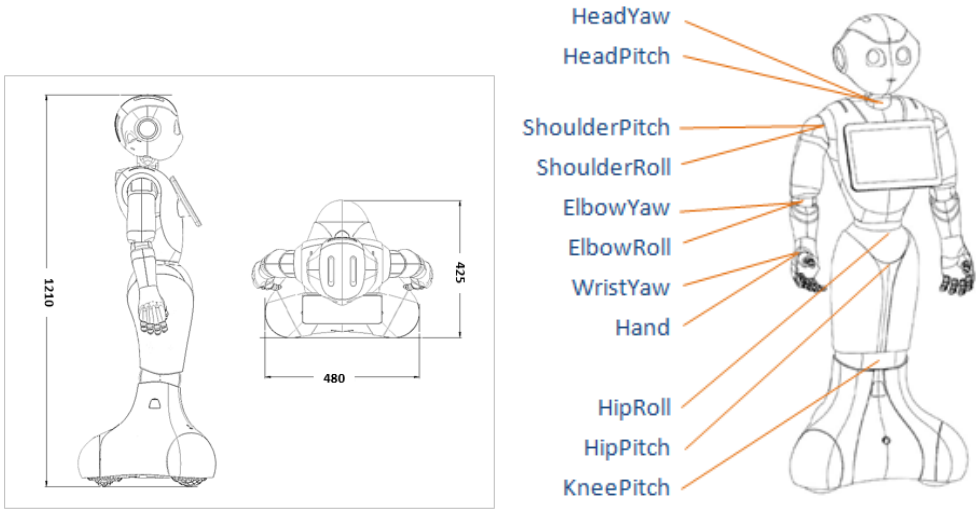
IMUs on unmanned autonomous vehicles are commonly used for navigation and calibration purposes. The cost of IMUs (in USD) is above 50k (used for navigation purposes), about 10-20k (used in tactical planning), 0.5-3k (used in industry), and below 0.5k (as a hobby) [74]. The price range of IMU devices is enormous; based on these devices' accuracy, the price fluctuates too much. In nonlinear estimation problems, extended Kalman filters are commonly used. It recursively estimates the system states that are corrupted with Gaussian noises from system measurements. Since gyroscope and accelerometer sensors have Gaussian-like noises, extended Kalman filters are an excellent option to be used with IMUs.

### 3.2.4 Semi-Humanoid Robot Pepper

Pepper, a social semi-humanoid robot by SoftBank Robotics, designed for people with friendly conversations, and its tablet can be used for business and education. The robot integrates a range of sensors for proximity and vision that allow tracking, localization, and navigation algorithms to be developed. The holonomic wheel drive system of the robot allows a broader range of movements that suit adequately in human-related scenarios. Pepper weighs about 29 kg, with much of this weight distributed in the lower part of the body. Notably, this design specification optimizes the upper body's movement, allowing Pepper to have a broader range of movements in his arms and torso (Fig: 3.7). The Pepper robot has many sensors attached to it, such as; three lasers located at the bottom of the robot, two sonars in the front and back of the KneePitch, two stereo cameras at the Pepper's head and mouth, an ASUS Xtion RGB-D camera located at its eyes, two Infra-Red sensors, 30 Magnetic Rotary Encoders (MRE), an Inertial Unit, four microphones, two loudspeakers, and LEDs at its eyes, ear, and shoulders. The field of view of laser, sonar, and cameras are shown in Fig: 3.8.

#### 3.2.4.1 NAOqi

NAOqi is the software that runs on the NAO and Pepper robots, and through NAOqi Framework, the robot can be controlled using Python, C++, and Java programming



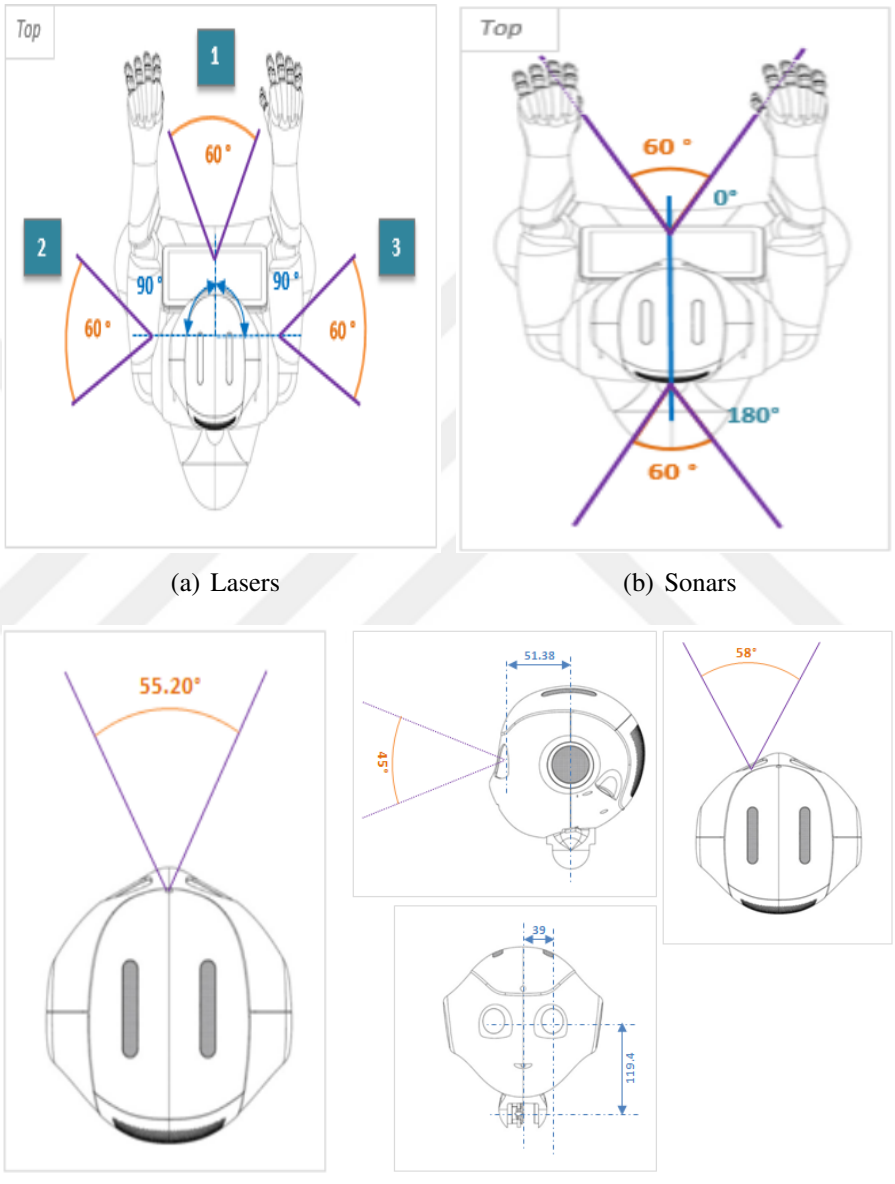

(a) Dimensions (b) Joints

**Figure 3.7** : Pepper robot's physical characteristics [75]

languages. NAOqi provides modules that are used to develop applications. NAOqi supports parallelism, synchronization, and events to the basic robotics needs. It covers all communication types between different modules and parts like motion, vision, microphone. NAOqi framework can cross many platforms such as Windows, Linux, or Mac and support identical API (Application Programming Interface) for both C++ and Python to benefit developers. The NAOqi has modules for motion, audio, people perception, event, and graphical user interface. These modules provide many APIs for definite services.

#### 3.2.4.2 Choregraphe

Choregraphe is a module of NAOqi to program Nao and Pepper robots in a graphical environment. It affords easy and efficient management of movements that enables users to develop moves rapidly. Users may use event-based, sequential, or parallel programming to design their programs. It also offers a timeline that helps users to construct their programs within the logic of the schedule. It can run on several platforms such as Windows, Linux, and Mac; moreover, it can call modules developed separately in C++ or Python through boxes or scripts written in Python. It can create behaviors a software to control the robot, with predefined behavior boxes or creating new boxes. Although Choregraphe has access to all the APIs of NAOqi, it is slower than behaviors written in C++.




(a) Lasers (b) Sonars

(c) Camera (d) Depth Camera

**Figure 3.8** : Pepper robot's field of view [75]

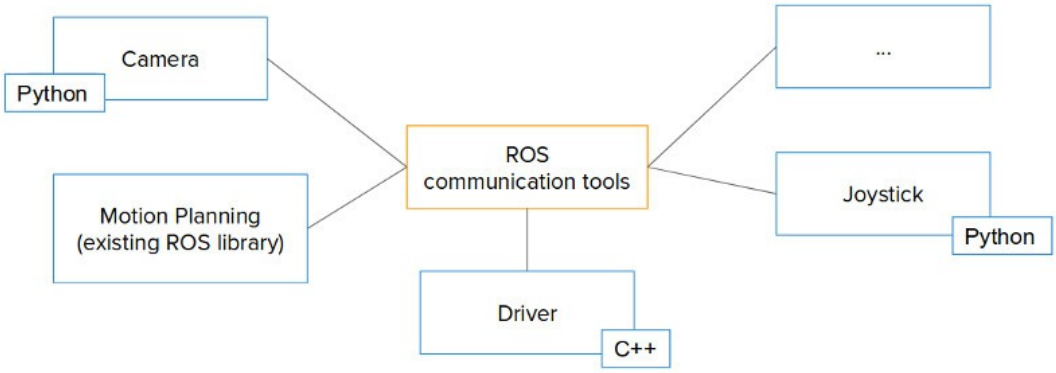



**Figure 3.9** : Language agnostic concept of ROS [76]

### 3.2.5 ROS

A robust structure for writing robot software is the Robot Operating System (ROS). It simplifies a wide range of robotic platforms to build complex and robust robot behavior. ROS was initially designed to facilitate collaborative robotics software development. It aims to provide robotic applications with a standard and use it on each robot without rewriting the codes for a different robot.

The concept of ROS is having independent blocks of codes that communicate through ROS communication tools. Each block of code can be responsible for different tasks such as camera, motion planning, and drivers. Since ROS is language-agnostic, blocks of codes can be in various programming languages like Python or C++ (Fig: 3.9). It currently supports Unix-based platforms; it is commonly used on Ubuntu systems.

Nodes, Master, Parameter Server, messages, services, topics, and bags are the basic Computation Graph concepts of ROS (Fig: 3.10) , all of which provide the Graph with data in various ways.

**Nodes** are computation-performing processes; one node controls the wheels, the other performs localization, and the system is composed of many nodes. **Master** provides name registration; without it, nodes would not find each other, exchange messages, or invoke services. The **Parameter** Server is a central location for storing data.

Nodes communicate by passing **messages**, a data structure with fields of primitive types such as integer and boolean and array of primitive types. The **topic** is like a message bus with a name; the messages are published by some nodes (publishers) and can be subscribed by some nodes (subscribers). The publishers and subscribers are not

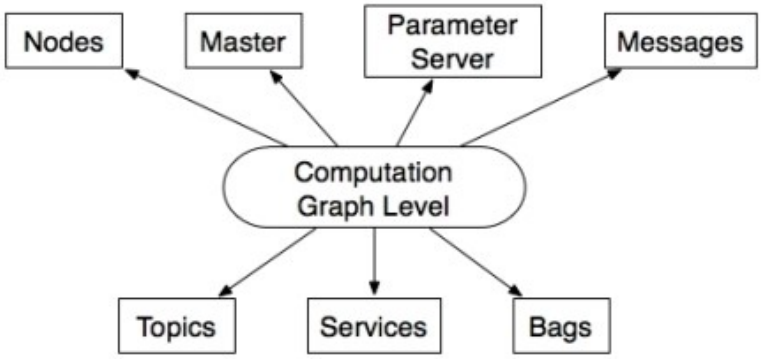



**Figure 3.10** : ROS communication concept [77]

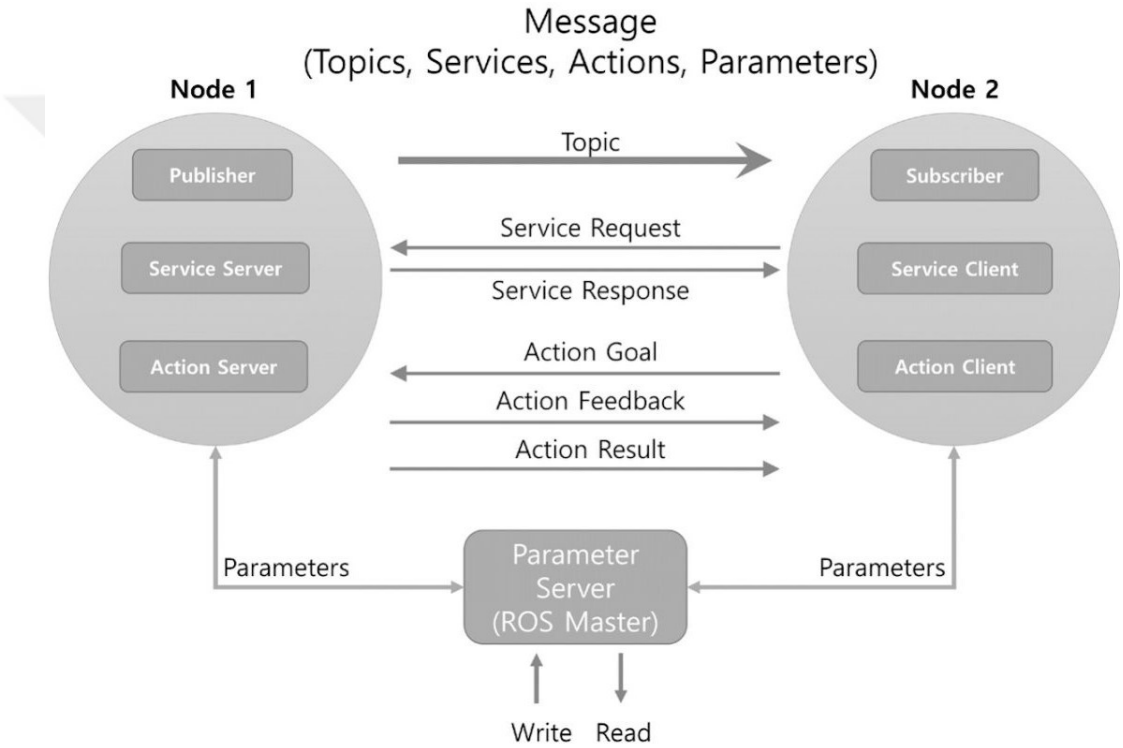



**Figure 3.11** : ROS communication between nodes [78]

aware of each other. **Services** are used for request/reply communication having two messages for request and reply; a node (server) provides the service, and other nodes (clients) request. When a client sends a request, it waits for a reply about the execution of the task. However, the user may want to cancel the request during execution or receive periodic feedback on how the request progresses if the service takes a long time to execute, and this is done through **actions**. The message communication is shown in Fig: 3.11 and Table 3.1 compares the topic, service and action with each other. **Bags** are a format for ROS message data to be saved and played back. Bags are an essential data storage mechanism, such as sensor data, that can be hard to gather but is crucial for algorithms to be developed and tested.

**Table 3.1** : Comparison of the Topic, Service and Action

| **Type** | **Features** | | **Usage** |
|---|---|---|---|
| Topic | Asynchronous | Unidirectional | Exchanging data continuously |
| Service | Synchronous | Bi-directional | Requesting processes and responding current states. |
| Action | Asynchronous | Bi-directional | When the service is hard to use due to long response times after the requestor when an intermediate value of the input is needed. |

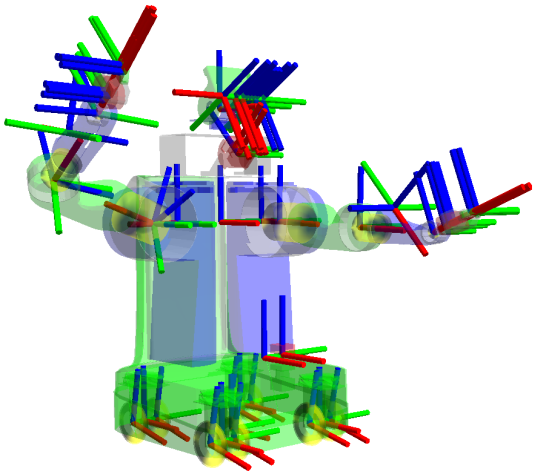
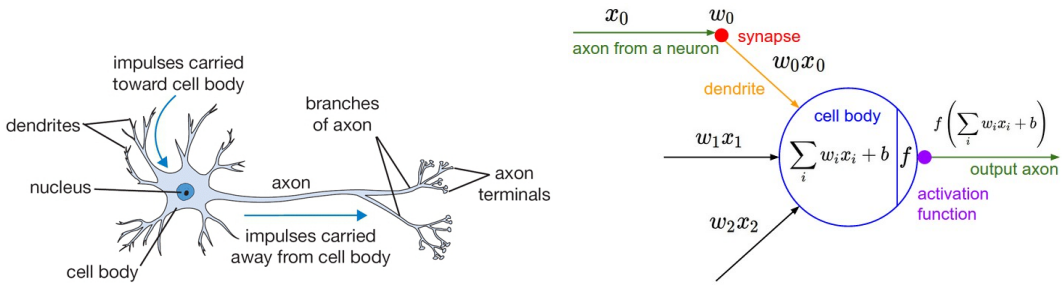


**Figure 3.12** : Coordinate frames of all components of a robot [79]

Usually, a robotic system has several 3D coordinate frames that change across time, such as a hand, head, gripper, and base frames. ROS transformation package TF retains the record of all these frames over time, allowing us to know at each moment the relative location of a frame to another frame. To know the robot's location in the real world, we need TF since it is computing each object's pose to itself, not the environment map. Furthermore, the objects are detected through cameras and sensors, but the robot's torso, gripper, and limb will contact the objects; thus, we need to know each object's relative pose to each robot's component that interacts or should avoid interaction with the object. As seen in Fig: 3.12 there can be many frames in a robot; ROS provides information about the robot's coordinate frames to all ROS components, which has simplified the robot's interaction with the environment.

#### 3.2.5.1 ROS-NAOqi Bridge

The Pepper robot is controlled with NAOqi framework, ROS-NAOqi package of the ROS provides links between ROS system components and NAOqi's APIs.




(a) Biological model (b) Computational model

**Figure 3.13** : Neuron [80]

## 3.3 Machine Learning Models

A mapping between test data entry and a set of output classes is generalized by Supervised Machine Learning. By analyzing a collection of training data with ground reality, the mapping is derived. This mapping is referred to as a predictive model, and the method of predicting a class label for a given data instance is classification.

### 3.3.1 Artificial Neural Network

Artificial Neural Network (ANN) is a computing system designed to simulate how information is analyzed and processed by the human brain (Fig: 3.13). In neuron cells, timing and frequency of signal matter; in the computational model, only frequency is considered with the activation function $f$ to transform the input signal to a new computational model.

ANNs have self-learning capabilities that allow them, as more data becomes available, to produce better results. Simple ANN has an output layer and an input layer. It is a linear classification with score function (maps raw data to class score) and loss function (quantifies the agreement between the predicted scores and the ground truth table). Thus it is a problem of optimization, minimizing the loss function with respect to the score function.

Different loss functions, such as Multiclass Support Vector Machine (SVM) classifier used Hing loss, and SoftMax classifier uses cross-entropy loss. SVM wants the correct class for each image to score higher than incorrect classes by a fixed margin $\Delta$. SoftMax classifier squashes class scores into normalized positive values that sum to

**Table 3.2** : SVM and Softmax classfiers

| Classifier | Loss function | |
|---|---|---|
| SVM | Hing loss | $L_i = \sum_{j \neq y_i} max(0, w_j^T x_i + \Delta)$ |
| Softmax | Cross entropy | $L_i = -\log(\frac{e^{f_{yi}}}{\sum_j e^{f_j}})$ |

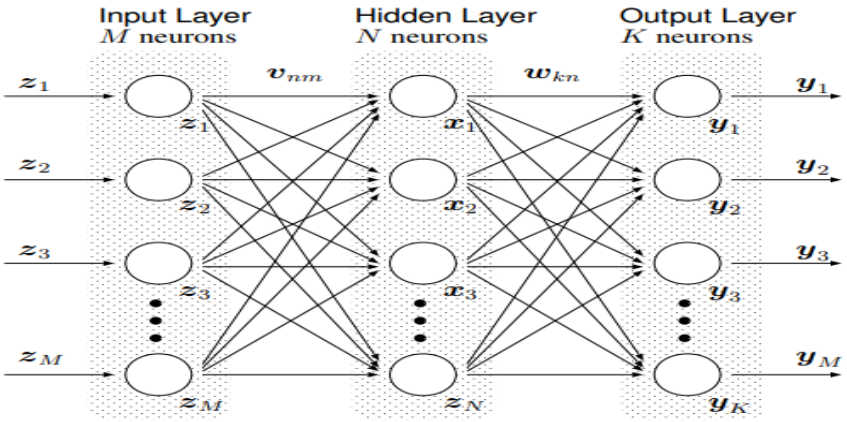



**Figure 3.14** : Multilayer perceptron (MLP) [81]

one so that the cross-entropy loss can be applied, then tries to minimize the correct class's negative log-likelihood as shown at Table 3.2

ANN with more than one layer, i.e., with the hidden layer(s), is called multilayer perceptron (MLP) (Fig: 3.14). The neurons in hidden layers of MLP have activation functions to break the linearity between the input and output layer. MLP performs a sequence of linear mapping with interwoven non-linearities. There are different activation functions, such as Sigmoid and Rectified Linear Unit (ReLU). ReLU is the recommended activation function because it is faster and computationally less expensive than other activation functions.

The **w** configuration is not inherently unique. Through an elementally quadratic penalty for all parameters, we prevent high weights. Penalizing helps to generalize better since no input dimension can all by itself have an enormous effect on the scores. This penalizing is done through regularization; ubiquitous regularizations are L1 and L2 (Table 3.3). L2 is usually preferred and it favors more diffuse weights.

Although the two-layer MLP is a universal approximator, MLPs with more than two layers theoretically have the same representational power but performs better in practice. The network's complexity also increases as the number of layers and the number of neurons in the layers increase, raising the overfitting problem.

**Table 3.3** : L1 and L2 regularizations

| Regularization | Equation | |
|---|---|---|
| L1 (lasso regression) | $R(W) = \sum_k \sum_n \lvert w_{k,n} \rvert$ | Estimating the median to avoid overfitting |
| L2 (ridge regression) | $R(W) = \sum_n \sum_n w_{k,n}^2$ | Estimating the mean to avoid overfitting |

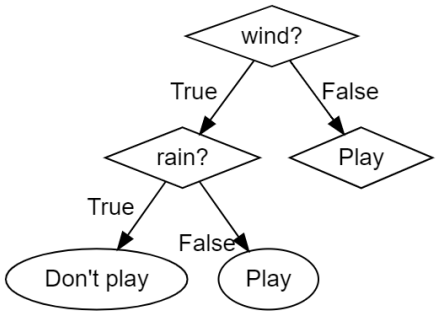



**Figure 3.15** : Decision tree example for the 'Play Golf' problem.

Regularization strength $\lambda$, not the number of neurons, is used to regulate the overfitting problem. Considering the regularization loss the total loss becomes as 3.2

$$L = \frac{1}{N} \sum_i L_i + \lambda R(W) \tag{3.2}$$

### 3.3.2 Decision Tree

It is a widely used and easy to understand machine learning model. It consists of internal nodes with a conditional test on a specific feature of the input data and guided edges dictating how these tests are performed. The final prediction class is found in Leaf nodes. A decision tree can be visualized with a node-like diagram. An example of the "Play Golf" problem is shown in Fig: 3.15.

The classification of a single instance occurs by following the route based on the node evaluation by following the root path down to the leaf. Training the model is a top-down process in which the training data set is divided into smaller subsets recursively. At each step, a feature is selected to split the data according to the model's splitting criterion; For each split set of data, a score is calculated, then the weighted average for all split sets is computed, which indicates the performance of the feature. Finally, the feature with the highest splitting score is selected for that node; this continues until no feature is left or not enough data points are left in that node or

no feature could result in a good splitting score according to the splitting criterion. The most common splitting criteria used in the training of decision trees are information gain and Gini impurity.

**Information gain**: This metric is based on information entropy, which calculates the data points' randomness. It favors small partitions with many distinct values. Formally, entropy $H$ is described by equation 3.3. Where $n$ is the number of classes, and $p_i$ refers to the likelihood of the class $Y_i$. $P_i$ is the fraction of class $Y_i$ within the subset S for decision tree classification. Information gain is defined as the difference between the entropy before the split and weighted average entropy after the split on feature $f$; it is shown in equation 3.4.

$$H(S) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{3.3}$$

$$G(S, f) = H(S) - H(S|f) \tag{3.4}$$

**Gini impurity**: It calculates the probability of misclassified classes when selected randomly. It favors large partitions. The equation is given in 3.5.

$$\begin{aligned} I_{Gini}(S) &= \sum_{i=1}^{n} p_i(1 - p_i) \\ &= 1 - \sum_{i=1}^{n} p_i^2 \end{aligned} \tag{3.5}$$

The well-known implementations of decision trees are ID3 and C4.5 using information gain and CART, which only constructs binary trees, uses Gini impurity.

### 3.3.3 Ensemble Models

New strategies have been proposed further to integrate the results of many independently trained models to boost classifiers' predictive accuracy.

**Bagging** works by sampling $M$ samples uniformly from the training dataset (sampling with replacement), where $N$ is the size of the original training dataset, and $M < N$. This is often referred to as a bootstrap sample [82]. A new classifier can be trained on any bootstrap sample. Among the number of classifiers, the final classification is decided by a majority vote.

**Boosting** classifiers are not independent as in Bagging; the model's training set is based on earlier classifiers' results. The probability of a sample being used as training data for the next model increases when a model wrongly classifies it. Therefore, Boosting seeks to build new classifiers that achieve greater accuracy for the records on which previous models performed poorly.

**Random forest** is the ensembles of decision trees with the Tree-Bagging method. In addition to Bagging, Random Forests (RF) also apply a method referred to as random subspace projection [83]. This method selects a random subset of features for each split candidate, which avoids overfitting and results in better generalization.

# 4. DATA COLLECTION and PREPROCESSING

In this work, the Myo armband device is used to capture sEMG and IMU signals. The armband sampling rate is reported 200 Hz and 50 Hz for EMG and IMU, respectively. In reality, Myo's sampling rates at the computer are below these values, and they depend on its Bluetooth range (device distance from the computer). Myo armband has eight dry sEMG electrodes and three inertial sensors for IMUs. It provides the inertial magnetometer sensor orientation in quaternion; thus, it provides 10 IMU values. The data of the device is reached through the Myo-connect application in multiple platforms such as Windows. This application can recognize by default six gestures (Fig: 3.1) by some extent and control PowerPoint and web browsers. Additionally, source code in both C++ and Python are available to access raw EMG and IMUs from the Myo-connect application.

At first, we wanted to train a classifier with available EMG datasets and see the performance of Myo in recognition of TSL sign gestures. The availalbe datasets are given in Table 4.1. The publicly available EMG datasets such as NinaPro(DB1), NinaPro(DB2), CapgMyo(DB-a), and Csl-hdemg have used different electrodes than the Myo armband, so they cannot be used in this work. The datasets from [49, 50, 84] have used the Myo armband to collect EMG signal for different hand gestures, but only one TSL gesture (V Hand) is among them. The publicly available datasets do not comply with this thesis's purpose because either the data is from different EMG electrodes than Myo armband, or the gestures do not contain TSL signs. Thus, we collected EMG and IMU datasets for TSL gestures.

## 4.1 Data Collection

We obtained EMG and IMUs data of ten healthy participants. Half of them were females of age 21-26 (average:23.8), and the other five were males of age 27-37 (average: 29.6). The detailed information is given in Table 4.2 The data is recorded for 36 static signs of TSL (Fig: 1.1 ) through Myo armband. Participants used their

**Table 4.1** : Available EMG datasets

| Dataset Name | Subjects | EMG Type | Number of Gesture |
|---|---|---|---|
| NinaPro(DB1) | 27 | 10 sEMG (Otto Bock) and CyberGlove II | 52 |
| NinaPro(DB2) | 40 | 12 sEMG (Delsys) and CyberGlove II | 50 |
| CapgMyo(DB-a) | 18 | 128 HD-sEMG | 8 |
| Csl-hdemg | 5 | 192 HD-sEMG | 27 |
| Myo-EMG [84] | 1 | 8 sEMG (Myo armband) | 8 |
| *Myo-EMG [49] | 35 | 8 sEMG (Myo armband) | 6 |
| *Myo-EMG [50] | 10 | 8 sEMG (Myo armband) | 6 |
| *Datasets are for gestures in Fig: 3.1 | | | |

**Table 4.2** : Participants in data collection

| Subject | Gender | Age | Dominant Hand |
|---|---|---|---|
| participant1 | Male | 27 | Right |
| participant2 | Female | 24 | Right |
| participant3 | Female | 23 | Left |
| participant4 | Male | 29 | Right |
| participant5 | Male | 33 | Right |
| participant6 | Female | 21 | Right |
| participant7 | Male | 32 | Right |
| participant8 | Male | 27 | Right |
| participant9 | Female | 25 | Right |
| participant10 | Female | 26 | Right |

dominant hand, where nine used their right hand except for one female who used her left hand. The gestures were performed without contracting the muscles with excess strength and moved their hand with all degrees of freedom (DOF) in the wrist, elbow, and shoulder joints (Table 4.3) to have data for a sign in various hand position. During data collection, each gesture was recorded separately, and the subject was given instruction about the gesture. The participants had not used the Myo armband before and they had no experience in sign language. If the gesture was recorded not correctly or any other error was happening; it was recorded again. Consequently, the average time to collect data from one participant was approximately 15 minutes.

For the GUI (graphical user interface), we used the *Tkinter* Python library; the GUI is shown in Fig: 4.1, where the *YeniKayit* opens a new window (Fig: 4.2) to add (name, gender, dominant hand, age, and history of muscle injury) info about the

**Table 4.3** : Degree of freedom (DOF) in wrist, elbow and shoulder joints

| Joint | DOF | |
|---|---|---|
| Wrist | 3 | (1) flexion/extension<br>(2) pronation/supination<br>(3) deviating ulnarly/radially |
| Elbow | 2 | (1) flexion/extension<br>(2) pronation/supination |
| Shoulder | 3 | (1) flexion/extension<br>(2) abduction<br>(3) rotation internal/external |

participant. From the left panel, we can select any participant who has been registered; thus, if a participant experienced muscle fatigue, he/she can halt the recording and proceed later. However, this did not happen with our participants. The *Baslat* button establishes/renew the connection with the Myo armband device. As the connection is established, the status will be shown at the right panel *MakeReady...* part. The right-panel shows the gesture to be performed. The *RekordEt* button vibrates the Myo armband, and after that, it starts recording; as the recording is finished, it gives the chance to accept the recorded data or drop it with *KabulEt* and *TekrarEt* buttons, respectively.

## 4.2 Preprocessing

The nature of sign gestures is more correlated with orientation than gyroscope and accelerometer; thus, we did not use any fusion algorithm (extended Kalman filter) on IMUs, which can perform better on gyroscope and accelerometer values. Furthermore, preprocessing the IMUs can not remarkably improve the categorization [85]. Thus we did not preprocess the IMU values, and we continued using quaternion values for orientation since they are numerically more stable and efficient in analyzing three-dimensional space than Euler angles [86].

The EMG signals are very noisy due to many factors, as explained in Section 3.2.2. Some of these noises, such as cross talk and internal noise, are better cleared inside the device because filters cannot efficiently block these noises. Thalmic labs have not provided information about noise cancellation inside Myo armband.

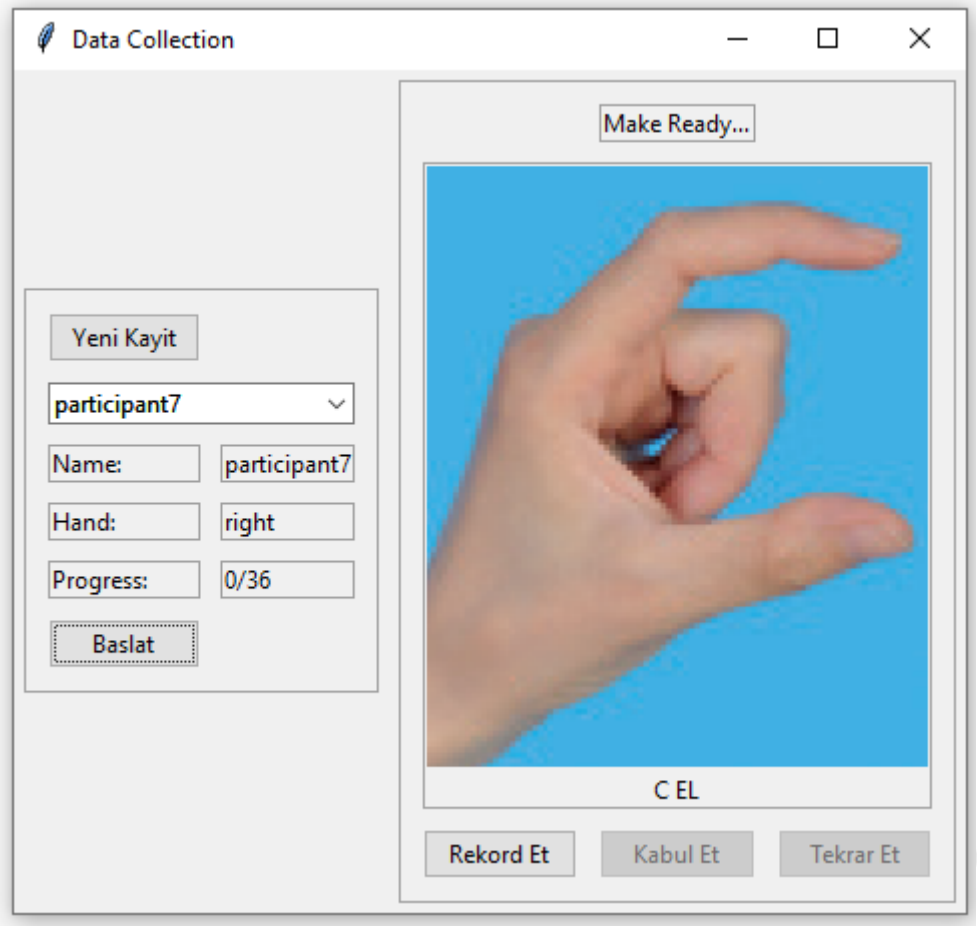



**Figure 4.1** : GUI for recording data

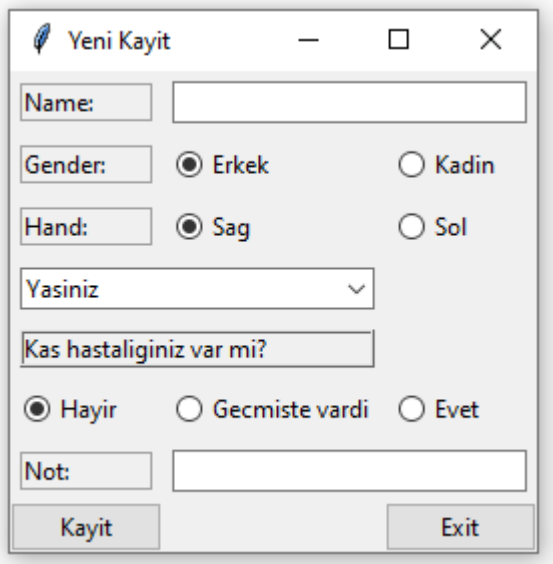



**Figure 4.2** : GUI for registering a new participant

A contraction becomes strong in two ways: increment the motor unit's firing rate and firing additional motor units. A regular firing rate of a motor unit is usually around 5 Hz; however, before another motor unit is added, the firing rate increases to 10 Hz. Considering the approximate frequency range of motor units, electromagnetic and electrocardiographic (ECG) noises can be filtered significantly with a low pass filter (allowing low-frequency signals only). The low pass filter usage can also block a wide range of noise resulting from inherent noise and movement artifacts.

To construct a digital filter, we need to transform the time-domain signal to frequency-domain and remove the signals with undesired frequencies (filtering); finally, transform the remaining signal back to the time-domain for further usage. Furthermore, speculating the signals at frequency-domain allows us to see each frequency's contribution to the combined signals' amplitude; this allows us to find the frequencies with less contribution (cut-off frequency), which indicates noise, to assign frequency range to our filters.

The Myo armband's EMG signal is quite noisy and enveloped between -100 and 100. The signal sign depends on whether the EMG sensor is near or far from the contracted muscle fibers; thus, both positive and negative values indicate the muscle contraction. We rectified the signal with the absolute-value function since muscles' action only depends on the signal's amplitude.

**Fourier transform (FT):** It is a mathematical transform that decomposes time-functions into frequency-functions. FT $\hat{f}$ is achieved as in Equation 4.1, and the inverse-FT $f$ (transforming frequency-function to time-function) is computed as in Equation 4.2. In these equations, $w$ and $t$ can be any real number indicating frequency and time, respectively, and the imaginary part can be transformed using the Euler formula (Equation 4.3).

$$\hat{f}(w) = \int_{-\infty}^{\infty} f(t) e^{-2j\pi tw} dt \tag{4.1}$$

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(w) e^{2j\pi tw} dw \tag{4.2}$$

$$e^{jx} = \cos x + j \sin x \tag{4.3}$$

The discrete Fourier transform (DFT) does the Fourier transform on a finite sequence of equally-spaced samples (Equation 4.4), and inverse-DFT (Equation 4.5) does the inverse-FT on the finite sequence. Where the parameters in the equations are; $T$ (number of samples), $t$ (current sample), $w$ (current frequency), $X_t$ (value of the signal at time $t$), and $X_w$ (amplitude and phase of frequency $w$). The $t$ and $w$ can only be zero or positive integer.

$$X_w = \sum_{t=0}^{T-1} X_t . e^{-2j\pi wt/T} \tag{4.4}$$

$$X_t = \frac{1}{T} \sum_{w=0}^{T-1} X_w . e^{-2j\pi wt/T} \tag{4.5}$$

Fast Fourier transform (FFT) is an algorithm that computes DFT and inverse-FFT performes inverse-DFT. It reduces the complexity of computing the DFT from $O(N2)$ (if DFT is simply computed) to $O(N \log N)$, where N is the data size.

After applying FFT on rectified sEMG data we found the cut-off frequency is below 5 Hz as illustrated in Fig: 4.3, for practical considerations we chose the cut-off frequency as 5 Hz. For filtering signals (blocking signals with frequencies higher than 5 Hz), we constructed a fourth-order, digital low pass Butterworth filter with a cut-off frequency of 5 Hz. The Butterworth is one of the best low pass filters available, and the 4th order is an excellent approximation to represent sin wave as quadratics. The digital filter is used in two directions, first to convert the signal to the frequency domain and, secondly, to transform the signal back to the time domain. As we feed a sample to the filter, we know the whole signal during processing; thus, there is no phase lag in this filtering. This whole filtering is achieved using the scientific python library (SciPy). The more detailed info about FT, DFT, FFT, and Butterworth filter is beyond the scope of this research. Fig: 4.3 shows the raw, rectified, filtered, and the Fourier transform for the rectified signal of sEMG.

### 4.3 Feature Extraction

Techniques for feature extraction, map the EMG into a feature set. These techniques extract characteristics in various domains, such as time, frequency, time-frequency, space, and fractal. Time-domain feature extraction is most common in EMG signals.

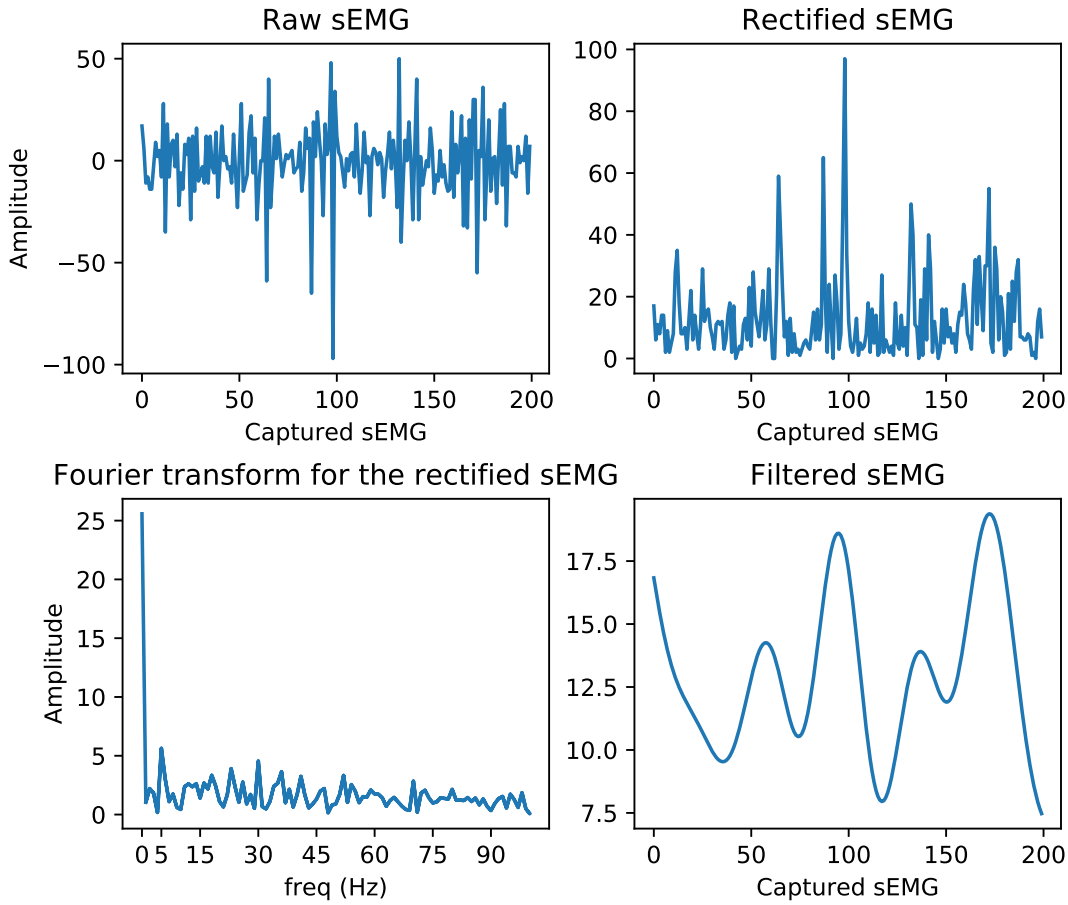



**Figure 4.3** : sEMG signal of ASL1 gesture for 1 second.

We wanted to see the effects of time-domain and time-frequency-domain features. We use preprocessed data (using frequency-domain) for extracting time-domain features, and since time-frequency-domain encapsulates both time and frequency domains; we do not use the preprocessed but the raw data.

We use the sliding window technique (Fig: 4.4) to extract features from the signals (dataset is separated into segments). For Myo EMG, the window length within [40, 80] EMG points is best for classification [22]. We tried multiple window lengths (40, 50, and 70). We choose the window length of 40 with the sliding length of 10; the 40 window length performed better than the other two lengths in classification. For Myo IMU values, we considered a window size of 10 (corresponds to 40 EMG points of Myo). Every window is treated and referred to as a sample from this point forwards.

### 4.3.1 Time-domain Features

A time-domain signal shows the changes in signal with time. The list of possible features that can be extracted from time-domain signals is long. Time-domain features

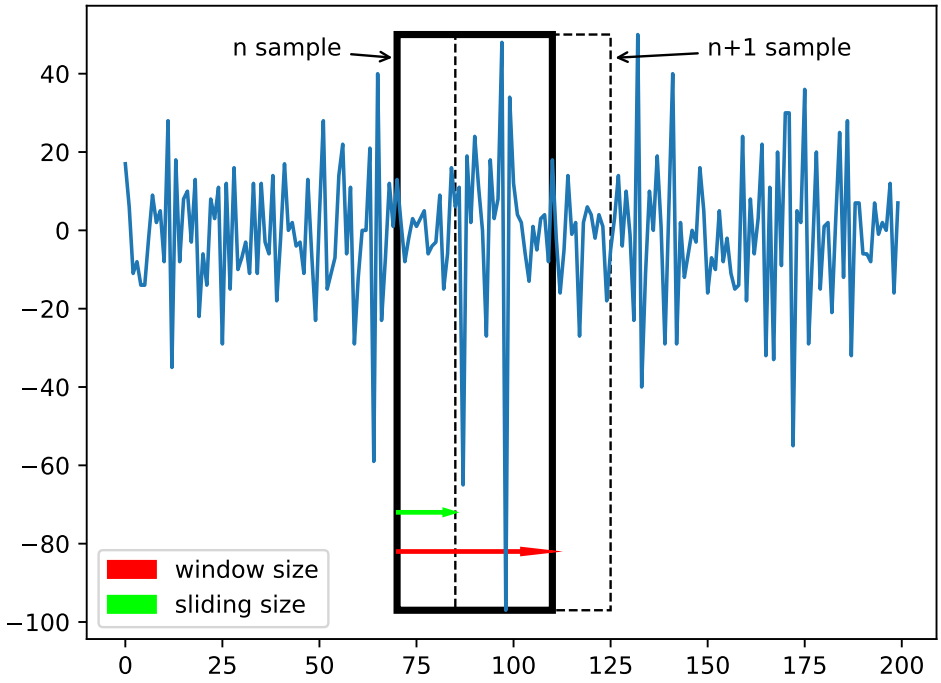



**Figure 4.4** : Sliding window technique used on sEMG signal

such as mean absolute value (MAV), root mean square (RMS), waveform length (WL), zero crossings (ZC), standard deviation (SD), variance (Var), slope sign changes (SSC), mean, median, integrated EMG (iEMG), sample entropy (SampEn), and many more have been extracted from EMG signals in multiple works.

For each IMU value, we extracted only one time-domain feature (mean). From the filtered EMG signal, five time-domain features are extracted that are common and important in analyzing EMG signals [58, 63]: waveform length, Hjorth parameters, mean, root mean square, and slope sign change.

1. Mean Absolute Value (MAV): It is the average of the rectified sEMG signal amplitude.

$$\text{MAV} = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{4.6}$$

2. Root Mean Square (RMS): It is the average power of the sEMG signal, which indicates the muscles' movement.

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2} \tag{4.7}$$

3. Slope Sign Change (SSC): This represents the number of slope sign variations in a sample.

$$\text{SSC} = \sum_{i=2}^{N-1} f(x_{i-1}, x_i, x_{i+1}), \text{ where}$$
$$f(x_{i-1}, x_i, x_{i+1}) = \begin{cases} 1, & \text{if } x_{i+1} > x_i > x_{i-1} \\ & \text{if } x_{i+1} < x_i < x_{i-1} \\ 0, & \text{otherwise} \end{cases} \tag{4.8}$$

4. Waveform Length (WL): The aggregate length of the sEMG signal waveform to consecutive points.

$$\text{WL} = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \tag{4.9}$$

5. Hjorth Parameters (HP): These are patterns of the signals' statistical features. These are three parameters: activity, mobility, and complexity. The parameter of activity describes the power of the signal, the variance of a time function. In the frequency domain, it can reflect the surface of the power spectrum.

$$\text{Activity}(f(x)) = var(f(x)) = \frac{1}{N-1} \sum_{i=1}^{N} (x_i)^2 \tag{4.10}$$

The mobility parameter reflects the mean frequency or the proportion of the standard deviation of the power spectrum.

$$\text{Mobility}(f(x)) = \sqrt{\frac{var(\frac{df(x)}{dx})}{var(f(x))}} \tag{4.11}$$

The complexity parameter compares the signal's similarity to a pure sine wave, where the value converges to 1 if the signal is more sine-like.

$$\text{Complexity}(f(x)) = \frac{\text{Mobility}(\frac{df(x)}{dx})}{\text{Mobility}(f(x))} \tag{4.12}$$

### 4.3.2 Time-frequency-domain Features

Studies a signal in both the time and frequency domains simultaneously. There are multiple possible ways to evaluate a signal in this way, such as short-time Fourier transform (STFT) and wavelet transform.

The FT provides information about the frequency (frequencies and their amplitudes) of a signal. It looks at the whole signal but does not provide information about a portion (localized region) of the signal in time. The STFT considers each localized

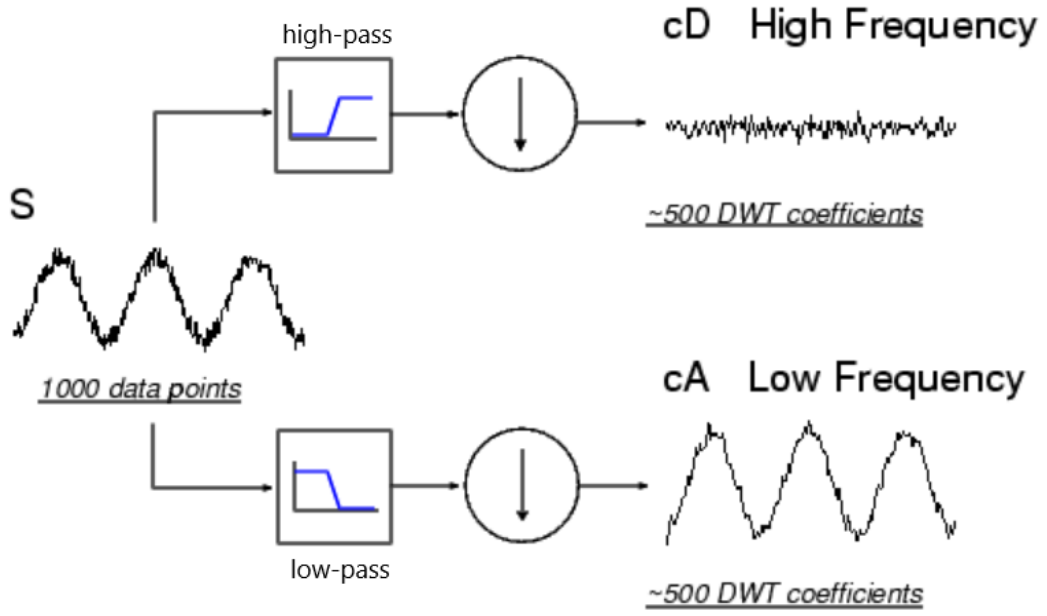
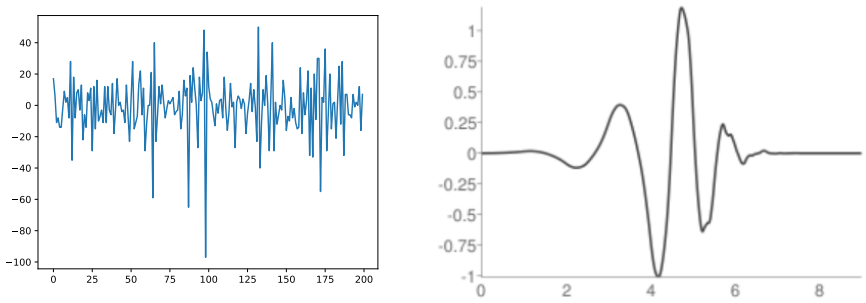



**Figure 4.5** : One-level DWT decomposition [87]

region of the signal separately and applies FT to that region; thus, it provides information about the signal at different times. The localized region is selected through a window function (windowing technique) with a fixed length. The window length affects the time and frequency resolutions as; narrow window (high time resolution, low frequency resolution) and wide window (low time resolution, high frequency resolution). However, low-frequency components of a signal usually continue for a long time (a wide window is required), and high-frequency components of a signal often appear for a shorter time (a narrow window is required). This problem is addressed with the wavelet transform; it uses the wavelet (Fig: 4.6(b); a wave-like oscillation that starts and ends with zero amplitude; the average value is also zero) as the window function at different scales. Calculating the wavelet coefficients at every possible scale generates many data; discrete wavelet transform (DWT) is used to eliminate this problem by choosing the scales on powers of two.

One level decomposition of DWT is passing the signal from both high-pass and low-pass filters in parallel. The decomposition doubles the signal's coefficients (same number of coefficients as the signal from both high-pass and low-pass filters); thus, we downsample the number of coefficients by averaging two consecutive coefficients (Fig: 4.5).

Generally, the low-frequency content of the signal is more critical than high-frequency. The decomposition process can be iterated, with successive approximations on




(a) Raw sEMG signal (b) db5 Wavelet [89]

**Figure 4.6** : sEMG signal and db5 wavelet

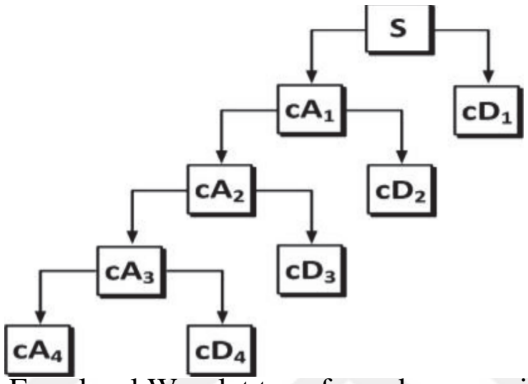



**Figure 4.7** : Four-level Wavelet transform decomposition tree [88]

low-pass filters' output (multiple-level decomposition). A family of orthogonal wavelets (Daubechies wavelets) based on Ingrid Daubechies' work are the most commonly used set of discrete wavelets used in DWT. We used four-layer Daubechies 5 wavelet, db5, to extract features (Fig: 4.6). Choosing 4-level db5 decomposition is borrowed from work [88]. The mathematics behind wavelet transform, DWT, Daubechies wavelets are beyond the scope of this work. The wavelet components of 4-level db5 decomposition tree (Fig: 4.7) are illustrated in Fig: 4.8.

The cA4, cD4, and cD3 were used with MAV and standard deviation (SD). Additionally, the MAV ratios between (cA4 and cD4, cD3, cD2) and (cD2 and cD3, cD4) have been extracted. We also used MAV and SD of the raw sEMG, while using only DWT as input to the classifier.

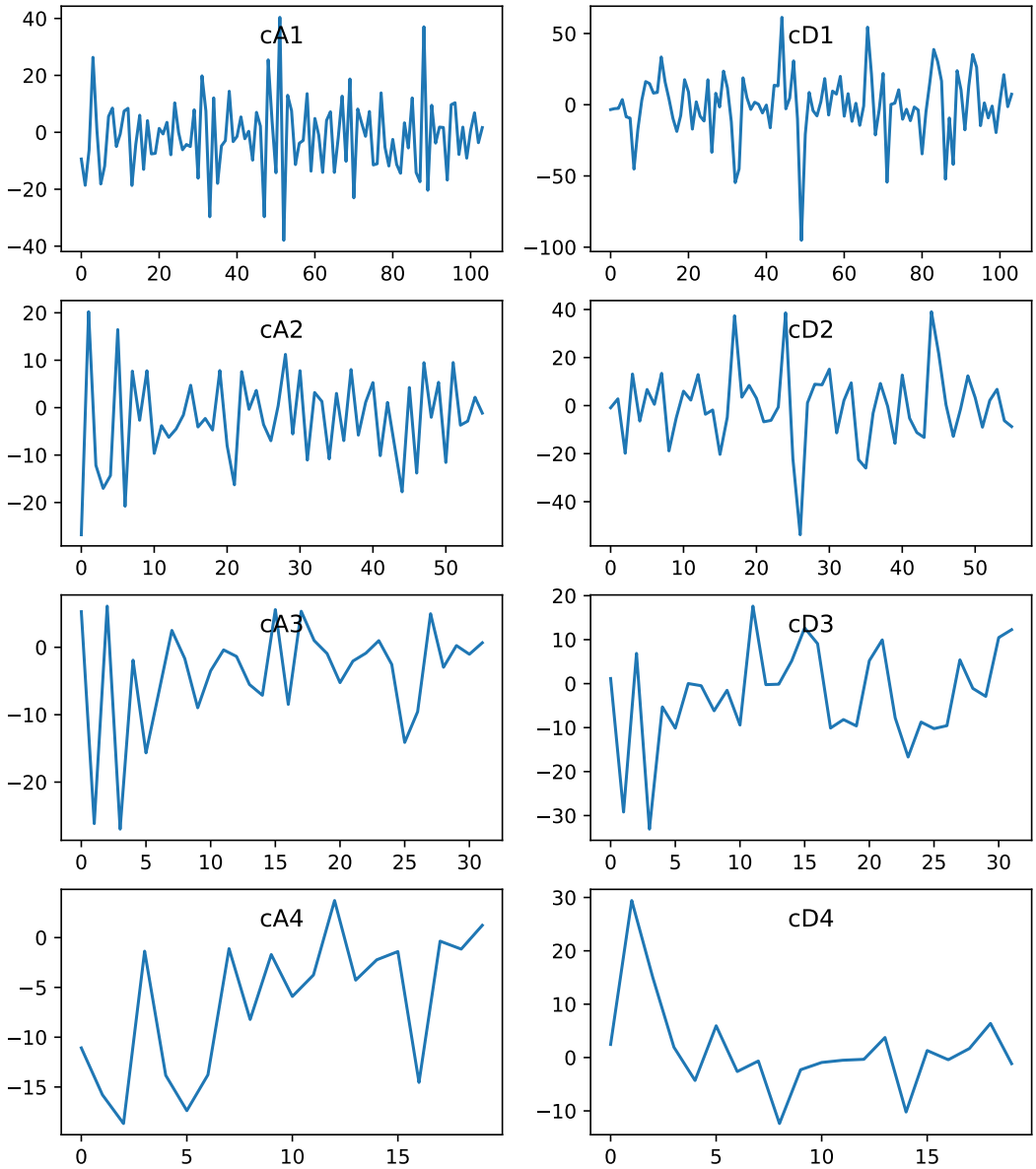



**Figure 4.8** : Four-level wavelet transform components

# 5. EXPERIMENTS AND RESULTS

We constructed a sign gesture classifier and controlled the Pepper robot's behavior with the recognized signs based on the proposed system. In this chapter, we review the finding from the TSL sign gesture recognition model. Then discuss the relations between gesture recognition, EMG, and IMUs. Finally, we explain the platform for controlling the robot's behavior.

## 5.1 Signs Recognition

Features are extracted from the electromyography signals in both time and time-frequency domains. This feature extraction is applied to the output of each electrode separately in addition to their cumulative sum. From this point forward, we refer time-domain and time-frequency-domain features to only sEMG features. The five time-domain features (where HP has three parameters) ends in 63 features (9x7). From the time-frequency-domain, we received eleven values that end in 99 features (9x11). Additionally, we have ten, averaged over a sample IMUs (three gyroscopes, three accelerometers, and four magnetometers values). Only the mean value is computed for each unit of the inertial sensors. There are three groups (time-domain, time-frequency-domain, and IMU) of features. Before training any classifier, we standardize the inputs to have an equal contribution from all features; therefore, we normalized the input data with zero mean and unit variance. In Fig: 5.1 the complete diagram for classifying gestures is presented.

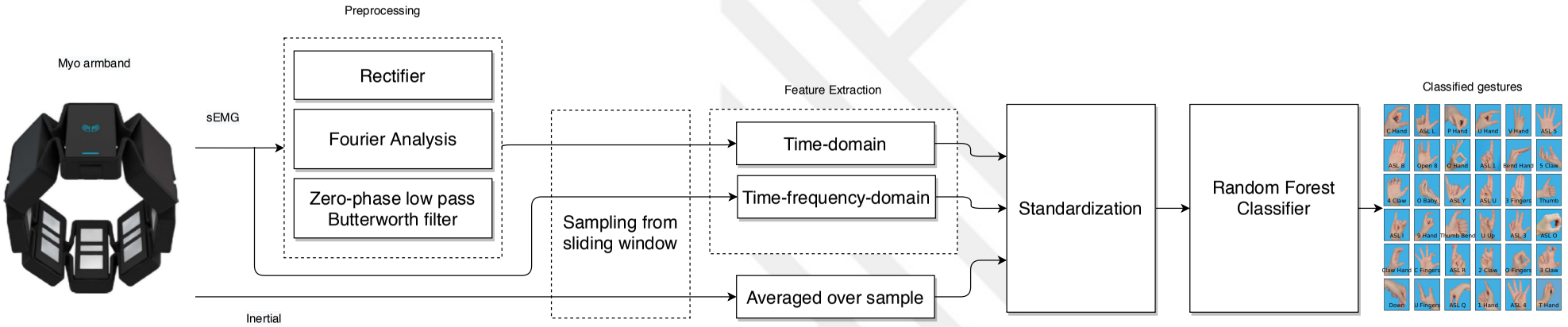



**Figure 5.1** : Complete diagram for classifying gestures

In this thesis, the accuracy metric is based on a cross-validation technique such that each time the data from one subject was fully considered the test set (leave-one-subject-out cross-validation). The MLP and RF classifiers are trained on the dataset. The non-trainable parameters (hyper-parameters) of MLP (such as the loss function, number of layers, and regularization strength) and RF (such as selection criterion in trees, number of trees, and depth of trees) are tunned with random selection over a wide range of possible combinations.

The accuracy of these classifiers was very low. Pretty much across all the signs, the RF classifier performed better in recognizing signs than MLP, so we advanced with RF. Table 5.1 presents the RF classifier's accuracy for various signs on one of the participants.

There are general and user-specific models. The general model is trained once, and then it can be used with new data from all users, including new users. However, the user-specific models are trained on one individual's data and then tested with the same user's new data. The general model is always preferred but is more difficult to achieve. Due to variation in data characteristics among different sources or if the model is used by one subject like hand prosthesis, a user-specific approach may be considered in the application.

When a gesture is performed, the envelope of electromyography signals varies for the same gesture performed by a different user. Additionally, the variation also exists between signals of two envelopes of a gesture performed by the same user [90]. Thus, constructing a general model that can be used by new users based on electromyography signals is very complicated. A general model based on a Myo armband can identify up to 7 gestures [91] with an acceptable accuracy; otherwise, the user-specific approach is required for a wide range of gestures.

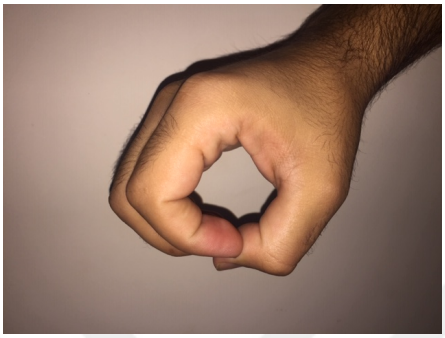
Our general model for all the static signs of TSL had a low accuracy score (Table 5.1), but it gave us the intuition that which signs are more capable of being recognized with Myo armband. Furthermore, to assess electromyography and inertial sensors' characteristics, we selected five reliable signs and proceeded with our research.

**Table 5.1** : Classification Accuracy for Different Gestures

| Gestures | Accuracy (%) |
|---|---|
| 1 Hand | 94 |
| Thumb Bend | 51 |
| ASL L | 41 |
| Thumb | 40 |
| ASL O | 38 |
| ASL 5 | 34 |
| ASL 4 | 21 |
| O Hand | 19 |
| C Fingers | 10 |
| Others | <10 |

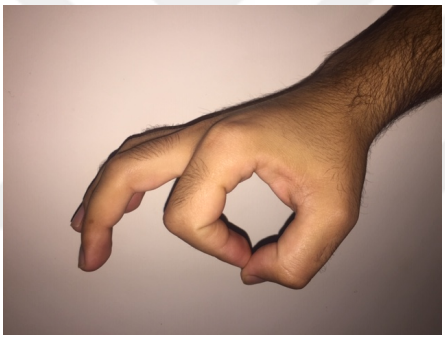
## 5.2 Selected TSL signs recognition

We selected five similar but more reliable sign gestures based on Table 5.1 (Fig: 5.2) to investigate the characteristics of EMG and IMUs further. We tunned and trained MLP and RF classifiers with different input sets (time-domain, time-frequency-domain, and IMUs). The RF classifier with 300 decision trees required a minimum of 5 samples to separate an internal node, maximum depth of 90, and bootstrap sampling performed better than any other classifier we trained. The highest classification accuracy, 54.2%, is achieved when all the extracted features from electromyography and inertial sensors were used. We trained several models based on gender, as well. Table 5.2 shows the classifier's accuracy with combinations of different groups of features.



(a) ASL O

(b) ASL L



(c) O Hand

(d) 1 Hand

(e) Thumb

**Figure 5.2** : Five selected TSL signs

**Table 5.2** : Classification Accuracy Using Different Sets of Features on Different Subjects

| Domain | #Features | Proportion of Variance (PoV) | Accuracy male (%) | Accuracy female (%) | Accuracy all (%) | Explanations |
|---|---|---|---|---|---|---|
| T | 63 | 1 | 74.0 | 22.2 | 51.3 | T: time-domain features |
| T | 15 | 0.993 | 73.4 | 22.4 | 50.9 | F: time-frequency-domain features |
| *F | 117 | 1 | 74.3 | 19.8 | 51.3 | I: inertial measurement units |
| *F | 20 | 0.993 | 73.3 | 19.9 | 51.3 | #Features: Number of inputs to the classifier |
| T & I | 73 | 1 | 73.0 | 25.8 | 50.5 | PoV: After the PCA applied the on the features |
| T & I | 20 | 0.995 | 56.2 | 24.1 | 41.6 | Accuracy male: Only male subjects were considered. |
| T & F | 162 | 1 | 78.0 | 18.2 | 51.1 | Accuracy female: Only female subjects were considered. |
| T & F | 20 | 0.992 | 28.7 | 24.0 | 25.2 | Accuracy all: All the subjects were considered. |
| T, F & I | 172 | 1 | 74.2 | 19.5 | 54.2 | *The MAV and SD of raw sEMG are also considered. |
| T,F & I | 20 | 0.992 | 28.9 | 23.6 | 26.3 | |

**Table 5.3** : Pepper robot's joints

| Joint | | Roll | Pitch | Yaw |
|---|---|---|---|---|
| Head | | x | ✓ | ✓ |
| Right Hand | Shoulder | ✓ | ✓ | x |
| | Elbow | ✓ | x | ✓ |
| | Wrist | x | x | ✓ |
| Left Hand | Shoulder | ✓ | ✓ | x |
| | Elbow | ✓ | x | ✓ |
| | Wrist | x | x | ✓ |

The machine learning models were better at recognizing the gestures performed by males than the female participants. To scrutinize the discrepancy in accuracy score based on gender, we plotted the gestures in the two-dimensional plane using t-distributed stochastic neighbor embedding (t-SNE) [92]. The output of t-SNE plots for female (Fig: 5.4) and male (Fig: 5.5) with both groups of features from time and time-frequency domains shows that the female gestures were more disperse than male gestures. We inferred that this is because the female subject had performed the signs with less muscular strength than the male. Therefore, training a machine learning classifier with reasonable accuracy on the female is more complicated.

Finally, after tuning the hyper-parameters of Random Forest and Multilayer perceptron models, they were trained solely on the male subjects for the five selected signs from TSL. They classify the signs correctly 78% and 68.3% of the time, respectively. The confusion matrix of the RF model is illustrated in Fig. 5.3. The signs ASL-O and ASL-L have generally performed with more muscular strength than their similar signs O-Hand and Thumb, respectively; thus, the model is less accurate in detecting the O-Hand and Thumb signs.

### 5.3 Simulating Pepper Robot on Gazebo

In this work, we want to control the hand position (head, shoulder, elbow and wrist joints) of the simulated semi-humanoid robot Pepper with the sign gestures recognized through Myo armband. Fig: 5.7 shows the axes defined on the Pepper robot, and the Table 5.3 have the information about all the joints of the Pepper robot that are controlled in this simulation.

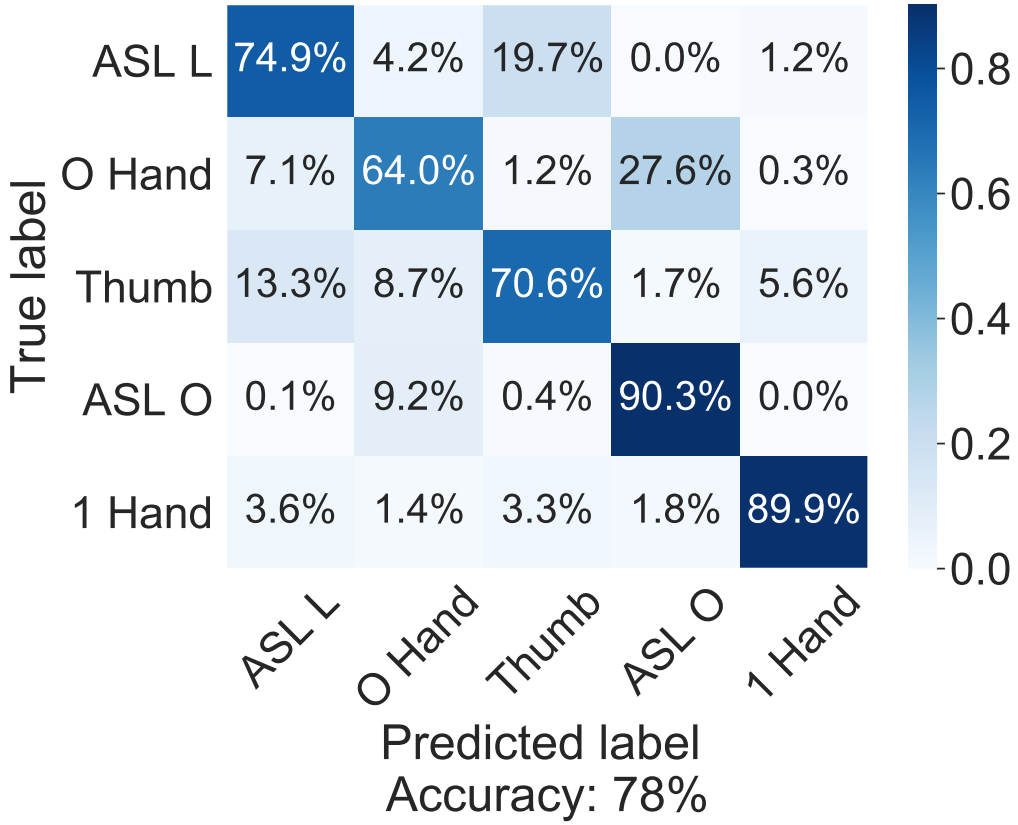



**Figure 5.3** : Normalized confusion matrix for five gestures.

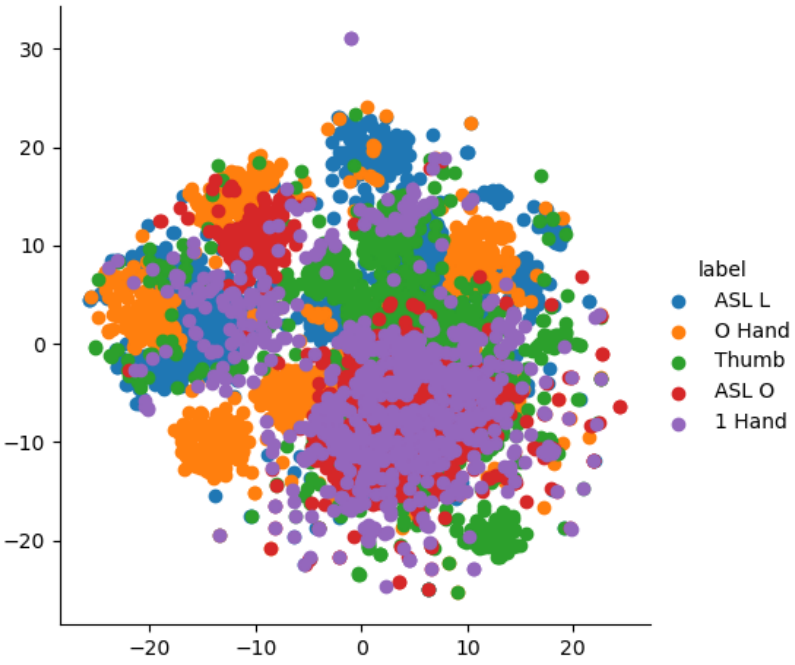



**Figure 5.4** : t-SNE of female subjects

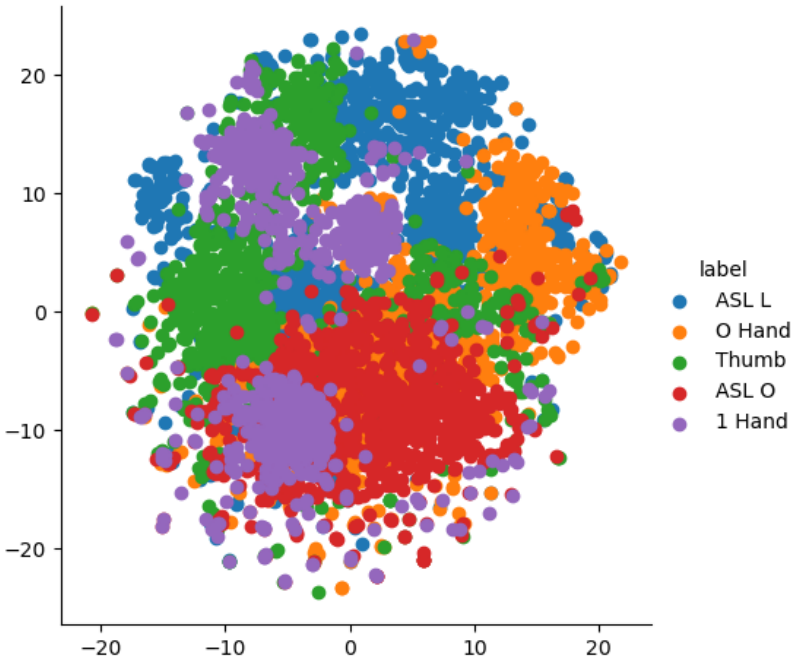



**Figure 5.5** : t-SNE of male subjects

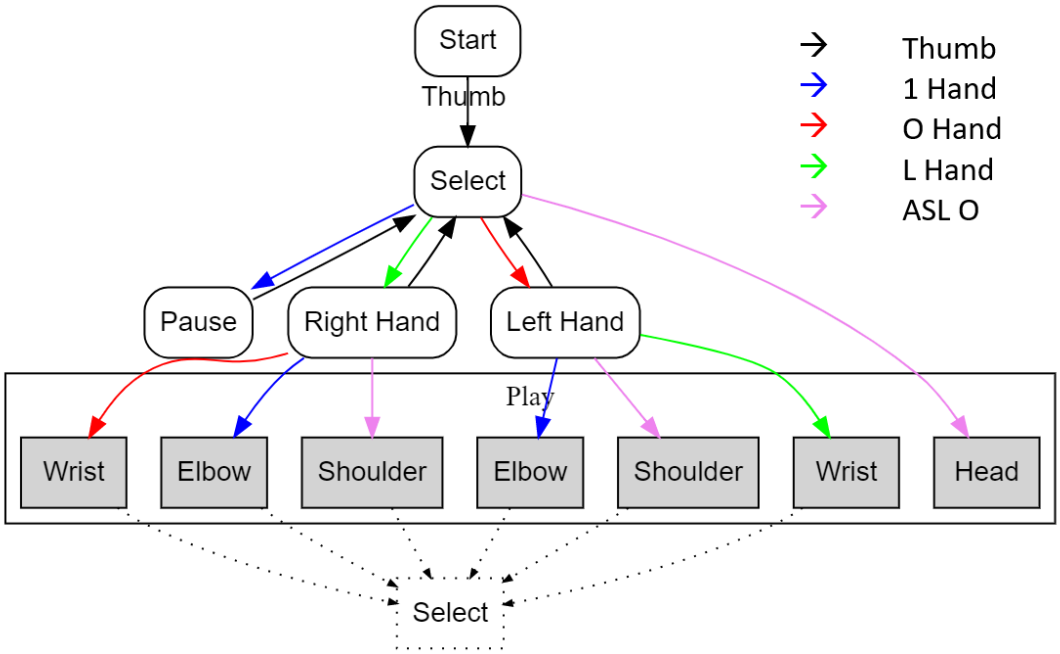



**Figure 5.6** : The state control of application

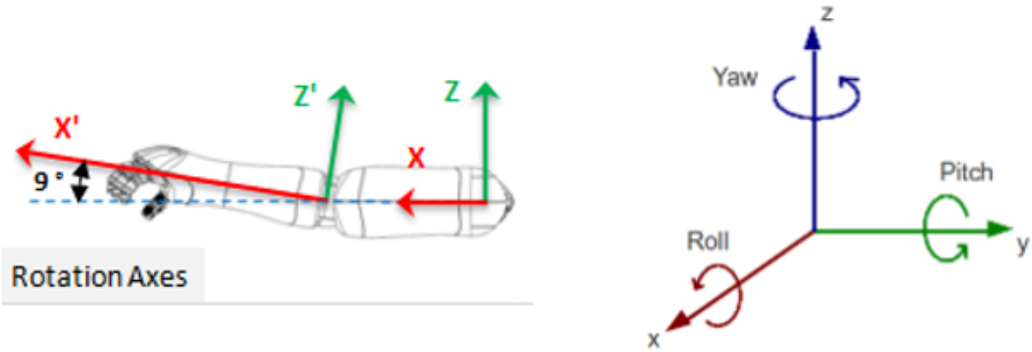



**Figure 5.7** : Rotation axes of Pepper robot [75]

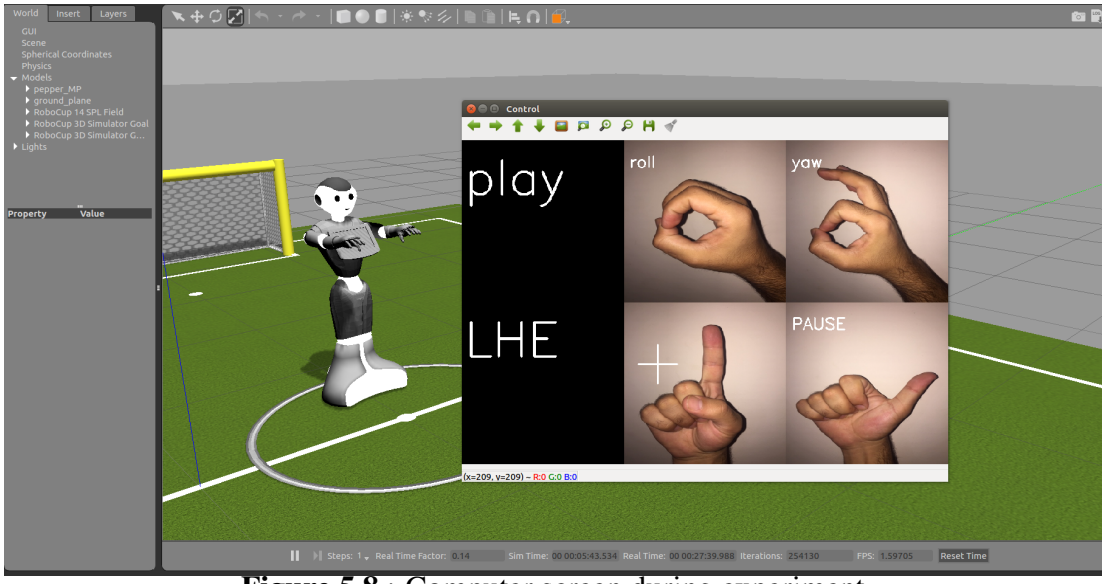



**Figure 5.8** : Computer screen during experiment

### 5.3.1 Experiment

We designed a GUI which is controlled through Myo armband. Each gesture is shown with a command; thus, as the user performs the gesture, the command is executed. To constraint, the undesired behavior postprocessing is applied to the gesture recognition model. The majority vote accepts the gesture in the last five results from the recognition model. Furthermore, the GUI allows the user to control 12 joints of the Pepper robot in both directions with five TSL gestures. A sequence of the gestures need to be recognized, GUI guides the user to select the control for a joint (Fig: 5.6). A specific gesture is needed to be performed to go to a state. No gesture can perform two tasks in a sequence; else, the user will lose full control over the program. As the control enters into a gray box, it asks whether to accept or reject the joint. As far as it is at state "Play," the sign gestures are shown with available rotations for that joint. A gesture is controlling the positive/negative direction of the joint. The screenshot of the computer during experiment is illustrated in Fig: 5.8 (GUI and Pepper robot on Gazebo). This experiment is based on Ubuntu 16.04 OS, ROS-Kinetics, and Pyo-connect (a Linux alternative to Myo-connect).

# 6. DISCUSSION

The EMG signal envelopes for the same hand gesture can have variations between and within users [90]. The reason for inter-class (between users) variance is the tissue structure and firing rate of motor units. Furthermore, the primary factors for EMG noise are skin formation, blood flow velocity, skin temperature, and tissue structure. These primary factors are all user-dependent, and they cannot be removed but reduced. Furthermore, the movement of sEMG electrodes on the skin results in movement artifact noise, which is also very user-dependent. The differences in humans' muscle structure and the noise factors in sEMG signals make the classification based on the sEMG signal more complicated to have a general model.

The research question about using EMG signal in the detection of TSL sign gestures can be addressed as; The TSL signs are very similar to each other with a little change in the fingers' position, a new sign is performed. Additionally, EMG signals have inter-class and intra-class variations for a gesture. These two issues make the detections of TSL signs with sEMG sensors challenging. It is possible to detect TSL signs with sEMG, but the accuracy and number of gestures are directly related to the electrodes' quality and number.

The accuracy of IMU values is directly proportional to the price of the sensor. The low-cost IMUs are around $500, and have above 1deg/sec error. The Myo armband costs $200; which primary components are as advertised are EMG electrodes. Although the accuracy comparison of Myo armband's IMU sensors, to the author's best knowledge, has not been reported in previous works, the cost and primary task of Myo armband can imply a considerable error in IMU values of this device. Furthermore, works that used Myo IMUs for controlling the screen (3x3 sections) have used a fixed pose to start recording, and at each section of the screen, the pose is reset based on the section coordinates [93, 94]. Thus, the error of IMUs can be reset after a short time, not to allow the error to exceed with time since only the current point is required to go to the next point. We cannot set fixed or reset locations because any sign can be

performed after any sign with any DOF at the wrist, elbow, and shoulder. Finally, we considered only static TSL sign gestures where gyroscope and accelerometers IMUs could not classify static signs in this work.

To address the second research question, using IMUs at detecting TSL signs, we can infer from the issues discussed earlier; The gyroscope and accelerometer IMUs can be used to recognize dynamic TSL signs but not static signs. The orientation IMUs can be efficient at classifying static TSL signs, but the sensor needs to be more accurate than the Myo armband's magnetometer inertial sensor.

The Myo armband having eight dry-sEMG and IMUs sensors is a low-cost commercially available device. The IMUs, especially orientation values of the armband, could be efficient in recognizing static TSL sign gestures but it turned out the Myo armband's magnetometer inertial sensor is very error-prone. Myo's sEMG electrodes are easy to use (without shaving and applying gel to the skin) but generate noisy EMG signals.

To address the research question, using the Myo armband for detecting TSL signs. The IMUs of Myo cannot significantly increase accuracy. The filtering and feature extraction of EMG signals of the Myo armband can be used to design classifiers for TSL signs. Due to noisy EMG signals and fewer electrodes in the Myo armband, it can classify a limited number of very different gestures that can be performed with high muscular strength (muscle contraction), and the gestures should not be performed with lots of DOF in the wrist, elbow, and shoulder. Thus having a general classifier with Myo armband for even a subset of TSL sign gestures is very complicated; Since the TSL sign gestures cannot be performed with high muscular strength, they are very similar to each other, and in our dataset, the gestures have had performed with any DOF at the wrist, elbow, and shoulder.

This study is a part of the ongoing RoboRehab (Assistive Audiology Rehabilitation Robot) project, where we aim to develop an effectively aware robotic rehabilitation platform to assist deaf children. In this project specific robot(s) is out of interest and different technologies such as Myo armband, Leap Motion, and other devices for sign recognition will work together on the same system. Although the Pepper robot's hand behaviors can be controlled easily with Choregraphe, this research aims to combine the

sign recognition from Myo armband to a standard robotic platform. With Choregraphe, we need to update our code if we want to use the NAO humanoid robot, and if we want to use any other robot, we have to rewrite the code again. Furthermore, with Choregraphe, we cannot directly use state-of-the-art algorithms for different tasks such as object recognition and path plannings that have not been explicitly developed for Pepper or NAO robots. Thus considering our project's aim, this research uses ROS to control the Pepper robot's behavior.

The Pepper robot has three versions 1.8, 1.8a, and 1.6. The drivers of these versions are different; the Pepper robot at our laboratory is version 1.6, and it requires manual calibration and modification to carefully control the robot. Furthermore, due to the Covid-19 pandemic, access to the laboratory, especially with many researchers at the same time, was difficult. Thus in this work, we decided to simulate the robot at Gazebo.

Gazebo is a robotics simulator, which can create 3D scenarios with robots, obstacles, and many other objects. It has physical engines (so that objects act like in the real world) for illumination, gravity, inertia, and more. It is designed to evaluate robot algorithms. ROS serves as a robot interface, and Gazebo sends information (the robot's perceived sensor data from the environment, joints status of the robot, and other information) to ROS. The ROS cannot differentiate between a simulation and a real robot. With ROS, we can run different algorithms for different tasks independently; based on the data it receives, we can do tasks such as behavior planning, object recognition, simultaneous localization and mapping, and path planning. Finally, we can send a command to the robot/simulation with ROS to control the robot. Gazebo gets the command and simulates the robot as if a real robot got that command in the real world.

Although ROS is a powerful tool to do sophisticated tasks with a robot, it works based on the robot or simulation program's sensor data. To simulate a robot in a scenario, Gazebo requires the description of the robot and its sensors. These descriptions need to be in a URDF (Unified Robot Description Format) file. The lasers are not defined in the robot's URDF file, and the RGB-D cameras of the simulated robot were not sending data to the ROS topic. The RGB-D issue could have been due to the virtual machine, Gazebo, or the operating system configurations. Since we could not receive

the data from the environment, we could not do localization, mapping, path planning, or interacting with other objects in the environment.

To address the research questions on integrating the sign recognizer model with robots; Some robots, especially humanoid robots like Pepper and NAO, have their graphical interface for defining robots' behavior, but it is better to use the ROS interface if we plan to do multiple tasks with the robot. Furthermore, we can reuse our code on a new robot by defining the communications between nodes without rewriting the code. Finally, ROS can save all the message data and play it back; this is crucial in developing and testing algorithms.

# 7. CONCLUSIONS AND FUTURE WORK

This thesis focuses on classifying stationary hand gestures from TSL using EMG and inertial sensors. This work's base motivation is to develop a sign recognizer model for Human-Robot Interaction to ease social skills learning in deaf children. The EMG and IMU signals are collected from the Myo armband. Because of the TSL signs' complexity, the Myo armband was not sufficient to recognize them all.

The TSL sign gestures are very similar to each other, with a little change in the fingers' position, a new sign is performed. Additionally, EMG signals have inter-class and intra-class variations for a gesture. These two issues make the detections of TSL signs with sEMG sensors challenging. It is possible to detect TSL signs with sEMG, but the accuracy and number of gestures are directly related to the electrodes' quality and number. Additionally, the gyroscope and accelerometer IMUs can be used to recognize dynamic TSL signs but not static signs. The orientation IMUs can be efficient at classifying static TSL signs, but the sensor needs to be more accurate than the Myo armband's magnetometer inertial sensor.

The filtering and feature extraction of EMG signals of the Myo armband can be used to design classifiers for TSL signs. Due to noisy EMG signals and fewer electrodes in the Myo armband, it can classify a limited number of very different gestures that can be performed with high muscular strength (muscle contraction), and the gestures should not be performed with lots of DOF in the wrist, elbow, and shoulder.

Besides the time-domain EMG features, the extracted features from the time-frequency-domain EMG affect the classifier's accuracy for better. The transformation from time to time-frequency is done with DWT, generally using Daubechies wavelets. The resulting DWT component can extract an extensive range of features; this can be very useful in complex deep convolutional neural networks.

There is inconsistency about testing and training protocols among the studies done on recognition of gestures using EMG. Some researchers have not specified those

protocols, where the comparison of related works becomes even more complicated. Therefore the reproducibility of their works is questionable [58]. Opposed to this thesis, other related researches such as [51], and [53] have proposed models specific to one user. Similarly, in comparison to this study, the researches in [50], [22], and [52] have proposed general models, but accuracy is reported on the same users that the classifiers have had their data during the training session.

The testing protocol in work [49] is similar to this thesis's protocol (cross-validation leave-one-subject-out); they have trained a classifier to distinguish six different hand gestures and recorded a 78% accuracy score. They have got a better accuracy because they have trained the model on considerably distinctive default gestures of Myo armband(Fig: 3.1). Furthermore, they have collected data from 35 subjects in laboratory-controlled conditions (greater muscular strength was applied with restriction in DOFs). We have introduced a model that can classify five TSL gestures that are subtle and similar to each other with reasonable accuracy. However, the main focus was to evaluate EMG and IMU sensors' applicability in classifying TSL sign gestures, where the findings from this thesis are a successful step for future research in this field.

The recommended model does not adequately perform well on female subjects, probably because the female participants were performing the gestures with less muscular strength. Thus, the EMG signals from their hand gestures were not distinguishable enough to train a classifier on them with acceptable accuracy.

The main contributions of this thesis are the followings:

1. The dataset for 36 stationary TSL hand gestures is created from ten healthy subjects.
2. The proposed method is a kind of gesture recognition that is capable of recognizing subtle TSL gestures.
3. The performance of extracted features from time and time-frequency domains of EMG signals and IMUs is evaluated in the classification of gestures.
4. An HRI platform to control a simulated robot in Gazebo through Myo armband is designed.

The obtained results from this work motivate us to develop our model and the HRI platform further. In future works, we will try to overcome the gender problem to enable more users to use the model in the HRI interface successfully. We will also develop the robotic platform to include more robots and more technologies such as Leap Motion and experiment with the real robot(s). Finally, we will evaluate the success rate of physical robots in the learning processes of children.

## REFERENCES


[1] **Cheok, M.J.**, **Omar, Z. and Jaward, M.H.** (2019). A review of hand gesture and sign language recognition techniques, *International Journal of Machine Learning and Cybernetics*, *10*(1), 131–153.

[2] **Kose, H.**, **Uluer, P.**, **Akalin, N.**, **Yorganci, R.**, **Özkul, A. and Ince, G.** (2015). The Effect of Embodiment in Sign Language Tutoring with Assistive Humanoid Robots, *International Journal of Social Robotics*, *7*, 537–548.

[3] **Gürpınar, C.**, **Uluer, P.**, **Akalın, N. and Köse, H.** (2019). Sign Recognition System for an Assistive Robot Sign Tutor for Children, *International Journal of Social Robotics*, 1–15.

[4] **Uluer, P.**, **Kose, H.**, **Oz, B.K.**, **Aydinalev, T.C. and Barkana, D.E.** (2020). Towards An Affective Robot Companion for Audiology Rehabilitation: How Does Pepper Feel Today?, *RO-MAN 2020-The 29th IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, pp.567–572.

[5] **Kivrak, H.**, **Uluer, P.**, **Kose, H.**, **Gumuslu, E.**, **Barkana, D.E.**, **Cakmak, F.**, **and Yavuz, S.** (2020). Physiological Data-Based Evaluation of a Social Robot Navigation System, *RO-MAN 2020-The 29th IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, pp.994–999.

[6] **Brim Jr, O.G.** (1968). Socialization through the life cycle.

[7] **Mayberry, R.**, **Segalowitz, S. and Rapin, I.**, (2002), Handbook of neuropsychology.

[8] **Bellugi, U. and Fischer, S.** (1972). A comparison of sign language and spoken language, *Cognition*, *1*(2-3), 173–200.

[9] **Yang, R.**, **Sarkar, S. and Loeding, B.** (2010). Handling Movement Epenthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 462–477.

[10] **Ying Wu and Huang, T.S.** (1999). Human hand modeling, analysis and animation in the context of HCI, *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, volume 3, pp.6–10 vol.3.

[11] **Zeshan, U.** (2003). Aspects of Türk Isaret Dili (Turkish Sign Language), *Sign Language & Linguistics*, *6*(1), 43–75.

[12] **Miles, M.** (2000). Signing in the Seraglio: mutes, dwarfs and jestures at the Ottoman Court 1500-1700, *Disability & Society*, *15*(1), 115–134.

[13] **Mehdi, S.A. and Khan, Y.N.** (2002). Sign language recognition using sensor gloves, *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02.*, volume 5, pp.2204–2206 vol.5.

[14] **Kubuş, O.** (2008). An analysis of Turkish Sign Language (TİD) phonology and morphology, *Ankara: METU: MA Thesis*.

[15] **Özel Eğitim ve Rehberlik Hizmetleri Genel Müdürlüğü (MEB)**, (2015 (accessed July 19, 2020)). Türk İşaret Dili Sözlüğü, `http://orgm.meb.gov.tr/alt_sayfalar/duyurular/1.pdf`.

[16] **Wu, Y.**, **Hu, X.**, **Wang, Z.**, **Wen, J.**, **Kan, J. and Li, W.** (2019). Exploration of Feature Extraction Methods and Dimension for sEMG Signal Classification, *Applied Sciences*, *9*(24), 5343.

[17] **Li, W.J.**, **Hsieh, C.Y.**, **Lin, L.F. and Chu, W.C.** (2017). Hand gesture recognition for post-stroke rehabilitation using leap motion, *2017 International Conference on Applied System Innovation (ICASI)*, IEEE, pp.386–388.

[18] **Islam, M.M.**, **Siddiqua, S. and Afnan, J.** (2017). Real time Hand Gesture Recognition using different algorithms based on American Sign Language, *2017 IEEE International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, pp.1–6.

[19] **Sarkar, A.**, **Patel, K.A.**, **Ram, R.G. and Capoor, G.K.** (2016). Gesture control of drone using a motion controller, *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*, IEEE, pp.1–5.

[20] **Pisharady, P. and Saerbeck, M.** (2015). Recent methods and databases in vision-based hand gesture recognition: A review, *Computer Vision and Image Understanding*, *141*, 152–165.

[21] **Jiménez, L.A.E.**, **Benalcázar, M.E. and Sotomayor, N.** (2017). Gesture recognition and machine learning applied to sign language translation, *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*, Springer, pp.233–236.

[22] **Zhang, Z.**, **Yang, K.**, **Qian, J. and Zhang, L.** (2019). Real-time surface emg pattern recognition for hand gestures based on an artificial neural network, *Sensors*, *19*(14), 3170.

[23] **Zhang, X.**, **Chen, X.**, **Wang, W.h.**, **Yang, J.h.**, **Lantz, V. and Wang, K.q.** (2009). Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors, *Proceedings of the 14th international conference on Intelligent user interfaces*, pp.401–406.

[24] **Khan, R.Z.**, **Ibraheem, N.A.**, **Meghanathan, N.** *et al.* (2012). Comparative study of hand gesture recognition system, *Proc. of International Conference of Advanced Computer Science & Information Technology in Computer Science & Information Technology (CS & IT)*, volume 2, pp.203–213.

[25] **Dipietro, L.**, **Sabatini, A. and Dario, P.** (2008). A Survey of Glove-Based Systems and Their Applications, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *38*, 461 – 482.

[26] **El-Sheimy, N.**, **Nassar, S. and Noureldin, A.** (2004). Wavelet de-noising for IMU alignment, *Aerospace and Electronic Systems Magazine, IEEE*, *19*, 32 – 39.

[27] **Yasen, M. and Jusoh, S.** (2019). A systematic review on hand gesture recognition techniques, challenges and applications, *PeerJ Computer Science*.

[28] **Darrell, T. and Pentland, A.** (1993). Space-time gestures, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp.335–340.

[29] **Braffort, A.**, **Bolot, L. and Segouat, J.** (2011). Virtual signer coarticulation in Octopus, a Sign Language generation platform, *GW 2011: The 9th International Gesture Workshop*.

[30] **Starner, T.**, **Weaver, J. and Pentland, A.** (1998). Real-time American sign language recognition using desk and wearable computer based video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(12), 1371–1375.

[31] **Shao, L.** (2016). Hand movement and gesture recognition using Leap Motion Controller, *Virtual Reality, Course Report*.

[32] **Lu, D.**, **Yu, Y. and Liu, H.** (2016). Gesture recognition using data glove: An extreme learning machine method, *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp.1349–1354.

[33] **Aran, O. and Akarun, L.** (2010). A multi-class classification strategy for Fisher scores: Application to signer independent sign language recognition, *Pattern Recognition*, *43*(5), 1776 – 1788.

[34] **Keskin, C. and Akarun, L.** (2009). Input-output HMM based 3D hand gesture recognition and spotting for generic applications (accepted for publication), *Pattern Recognit Lett*, *30*(12), 1086–1095.

[35] **Caplier, A.**, **Stillittano, S.**, **Aran, O.**, **Akarun, L.**, **Bailly, G.**, **Beautemps, D.**, **Aboutabit, N. and Burger, T.** (2008). Image and video for hearing impaired people, *EURASIP Journal on Image and Video Processing*, *2007*(1), 045641.

[36] **Demircioğlu, B.**, **Bülbül, G. and Köse, H.** (2016). Turkish Sign Language recognition with Leap Motion, *2016 24th Signal Processing and Communication Application Conference (SIU)*, pp.589–592.

[37] **Karaci, A.**, **Akyol, K. and Gültepe, Y.** (2018). Turkish Sign Language Alphabet Recognition with Leap Motion, *Proceedings of the International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES'18), Safranbolu, Turkey*, pp.11–13.

[38] **Kaya, F.**, **Tuncer, A.F. and Yildiz, S.K.** (2018). Detection of the turkish sign language alphabet with strain sensor based data glove, *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp.1–4.

[39] **Kakajan, K. and Songül, A.** (2016). TURKISH SIGN LANGUAGE RECOGNITION USING HIDDEN MARKOV MODEL, *CS & IT-CSCP*, 11–18.

[40] **Englehart, K. and Hudgins, B.** (2003). A robust, real-time control scheme for multifunction myoelectric control, *IEEE Transactions on Biomedical Engineering*, *50*(7), 848–854.

[41] **Stifani, N.** (2014). Motor neurons and the generation of spinal motor neurons diversity, *Frontiers in cellular neuroscience*, *8*, 293.

[42] **Singh, R.E.**, **Iqbal, K.**, **White, G. and Holtz, J.K.** (2019). A review of EMG techniques for detection of gait disorders, *Artificial Intelligence-Applications in Medicine and Biology*.

[43] **Krasoulis, A.**, **Kyranou, I.**, **Erden, M.S.**, **Nazarpour, K. and Vijayakumar, S.** (2017). Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements, *Journal of neuroengineering and rehabilitation*, *14*(1), 71.

[44] **Borisov, I.I.**, **Borisova, O.V.**, **Krivosheev, S.V.**, **Oleynik, R.V. and Reznikov, S.S.** (2017). Prototyping of EMG-controlled prosthetic hand with sensory system, *IFAC-PapersOnLine*, *50*(1), 16027–16031.

[45] **Wu, J.**, **Tian, Z.**, **Sun, L.**, **Estevez, L. and Jafari, R.** (2015). Real-time American sign language recognition using wrist-worn motion and surface EMG sensors, *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, IEEE, pp.1–6.

[46] **Cheng, J.**, **Chen, X.**, **Liu, A. and Peng, H.** (2015). A novel phonology-and radical-coded chinese sign language recognition framework using accelerometer and surface electromyography sensors, *Sensors*, *15*(9), 23303–23324.

[47] **Yang, C.**, **Long, J.**, **Urbin, M.A.**, **Feng, Y.**, **Song, G.**, **Weng, J. and Li, Z.** (2018). Real-Time Myocontrol of a Human–Computer Interface by Paretic Muscles After Stroke, *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 1126–1132.

[48] **Savur, C. and Sahin, F.** (2015). Real-Time American Sign Language Recognition System Using Surface EMG Signal, *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp.497–502.

[49] **Nasri, N.**, **Orts-Escolano, S.**, **Gomez-Donoso, F. and Cazorla, M.** (2019). Inferring static hand poses from a low-cost non-intrusive sEMG sensor, *Sensors*, *19*(2), 371.

[50] **Benalcázar, M.E.**, **Anchundia, C.E.**, **Zea, J.A.**, **Zambrano, P.**, **Jaramillo, A.G. and Segura, M.** (2018). Real-time hand gesture recognition based on artificial feed-forward neural networks and emg, *2018 26th European Signal Processing Conference (EUSIPCO)*, IEEE, pp.1492–1496.

[51] **Abreu, J.G.**, **Teixeira, J.M.**, **Figueiredo, L.S. and Teichrieb, V.** (2016). Evaluating sign language recognition using the myo armband, *2016 XVIII Symposium on Virtual and Augmented Reality (SVR)*, IEEE, pp.64–70.

[52] **Luh, G.C.**, **Ma, Y.H.**, **Yen, C.J. and Lin, H.A.** (2016). Muscle-gesture robot hand control based on sEMG signals with wavelet transform features and neural network classifier, *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 2, IEEE, pp.627–632.

[53] **Kaya, E. and Kumbasar, T.** (2018). Hand Gesture Recognition Systems with the Wearable Myo Armband, *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, IEEE, pp.1–6.

[54] **Vygotsky, L.S.** (1980). *Mind in society: The development of higher psychological processes*, Harvard university press.

[55] **Steve, M. and Aaron**, (2014 (accessed July 19, 2020)). Thalmic Labs. Big data: Raw emg free for developers in december, `http://developerblog.myo.com/big-data/`.

[56] **Visconti, P.**, **Gaetani, F.**, **Zappatore, G.**, **Primiceri, P.** *et al.* (2018). Technical features and functionalities of Myo armband: an overview on related literature and advanced applications of myoelectric armbands mainly focused on arm prostheses, *International Journal on Smart Sensing and Intelligent Systems*, *11*(1), 1–25.

[57] **Mendez, I.**, **Hansen, B.W.**, **Grabow, C.M.**, **Smedegaard, E.J.L.**, **Skogberg, N.B.**, **Uth, X.J.**, **Bruhn, A.**, **Geng, B. and Kamavuako, E.N.** (2017). Evaluation of the Myo armband for the classification of hand motions, *2017 International Conference on Rehabilitation Robotics (ICORR)*, pp.1211–1214.

[58] **Jaramillo, A.**, **Benalcázar, M. and Mena-Maldonado, E.** (2020). Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review, *Sensors*, *20*, 2467.

[59] **Raff, H.**, **Falk-Steinmetz, M.**, **Medhora, M.**, **De Roo, K. and Holt, P.** (2014). In Vander's Human Physiology: The Mechanisms of Body Function, *Undersea & hyperbaric medicine: journal of the Undersea and Hyperbaric Medical Society, Inc*, *41*(4), 331–5.

[60] **Weiss, L.**, **Weiss, J. and Silver, J.** (2015). *Easy EMG E-Book: A Guide to Performing Nerve Conduction Studies and Electromyography*, Elsevier

Health Sciences, https://books.google.com.tr/books?id=l1emBgAAQBAJ.

[61] **Leis, A.A. and Schenk, M.P.** (2013). *Atlas of nerve conduction studies and electromyography*, Oxford University Press.

[62] **Nishihara, K.**, **Chiba, Y.**, **Suzuki, Y.**, **Moriyama, H.**, **Kanemura, N.**, **Ito, T.**, **Takayanagi, K. and Gomi, T.** (2010). Effect of position of electrodes relative to the innervation zone onsurface EMG, *Journal of medical engineering & technology*, *34*(2), 141–147.

[63] **Chowdhury, R.H.**, **Reaz, M.B.**, **Ali, M.A.B.M.**, **Bakar, A.A.**, **Chellappan, K. and Chang, T.G.** (2013). Surface electromyography signal processing and classification techniques, *Sensors*, *13*(9), 12431–12466.

[64] **Garcia, M.C. and Vieira, T.** (2011). Surface electromyography: Why, when and how to use it, *Revista andaluza de medicina del deporte*, *4*(1), 17–28.

[65] **Webster, J.G.** (1990). *Encyclopedia of medical devices and instrumentation*, John Wiley & Sons, Inc.

[66] **Reaz, M.B.I.**, **Hussain, M.S. and Mohd-Yasin, F.** (2006). Techniques of EMG signal analysis: detection, processing, classification and applications, *Biological procedures online*, *8*(1), 11–35.

[67] **Viljoen, S.**, **Hanekom, T. and Farina, D.** (2007). Effect of characteristics of dynamic muscle contraction on crosstalk in surface electromyography recordings, *S. Afr. Inst. Electr. Eng*, *98*, 18–28.

[68] **Stoykov, N.**, **Lowery, M.**, **Taflove, A. and Kuiken, T.A.** (2001). A finite element analysis of muscle tissue capacitive effects and dispersion in EMG, *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, IEEE, pp.1044–1047.

[69] **Taelman, J.**, **Mijovic, B.**, **Van Huffel, S.**, **Devuyst, S. and Dutoit, T.** (2011). ECG Artifact Removal from Surface EMG Signals by Combining Empirical Mode Decomposition and Independent Component Analysis., *Biosignals*, pp.421–424.

[70] **Groves, P.** (2008). GNSS Technology and Aplications Series, *Principles of GNSS, Inertial and Multisensor Integrated Navigation Systems. London: Artech House*.

[71] **Kajánek, P.** (2014). Testing of the possibilities of using IMUs with different types of movements, *Geoinformatics FCE CTU*, *12*, 61–66.

[72] **Titterton, D.**, **Weston, J.L. and Weston, J.** (2004). *Strapdown inertial navigation technology*, volume 17, IET.

[73] **Farrell, J.** (2008). *Aided navigation: GPS with high rate sensors*, McGraw-Hill, Inc.

[74] **Chao, H.**, **Coopmans, C.**, **Di, L. and Chen, Y.** (2010). A comparative evaluation of low-cost IMUs for unmanned autonomous systems, *2010 IEEE Conference on Multisensor Fusion and Integration*, pp.211–216.

[75] **Url-2**, <*http://doc.aldebaran.com/2-4/family/pepper_technical/index_pep.html*>, date retrieved 07.01.2021.

[76] **Url-3**, <*https://roboticsbackend.com/what-is-ros/*>, date retrieved 07.01.2021.

[77] **Martinez, A. and Fernández, E.** (2013). *Learning ROS for robotics programming*, Packt Publishing Ltd.

[78] **Url-5**, <*https://robinrobotic.blogspot.com/2019/06/ros-terminology.html*>, date retrieved 07.01.2021.

[79] **Url-1**, <*http://wiki.ros.org/tf*>, date retrieved 19.01.2021.

[80] **Url-4**, <*https://cs231n.github.io/neural-networks-1*>, date retrieved 07.01.2021.

[81] **Isokawa, T.**, **Nishimura, H. and Matsui, N.** (2012). Quaternionic multilayer perceptron with local analyticity, *Information*, *3*(4), 756–770.

[82] **Efron, B. and Tibshirani, R.J.** (1994). *An introduction to the bootstrap*, CRC press.

[83] **Ho, T.K.** (2002). A data complexity analysis of comparative advantages of decision forest constructors, *Pattern Analysis & Applications*, *5*(2), 102–112.

[84] **Espinoza, D.L. and Velasco, L.E.S.** (2020). Comparison of EMG signal classification algorithms for the control of an upper limb prosthesis prototype, *2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, IEEE, pp.1–4.

[85] **Fida, B.**, **Bernabucci, I.**, **Bibbo, D.**, **Conforto, S. and Schmid, M.** (2015). Pre-Processing Effect on the Accuracy of Event-Based Activity Segmentation and Classification through Inertial Sensors, *Sensors*, *15*, 23095–23109.

[86] **Zomorodi, S. and Engineer, S.E.**, Quaternions Approach in Studying Rotation.

[87] **Url-7**, <*http://matlab.izmiran.ru/help/toolbox/wavelet/ch01_31a.html*>, date retrieved 07.01.2021.

[88] **Ohashi, H.**, **Al-Nasser, M.**, **Ahmed, S.**, **Akiyama, T.**, **Sato, T.**, **Nguyen, P.**, **Nakamura, K. and Dengel, A.** (2017). Augmenting wearable sensor data with physical constraint for DNN-based human-action recognition, *ICML 2017 Times Series Workshop*, pp.6–11.

[89] **Url-6**, <*http://wavelets.pybytes.com/wavelet/db5/*>, date retrieved 07.01.2021.

[90] **Rodriguez-Falces, J.**, **Navallas, J. and Malanda, A.** (2012). EMG modeling, *Computational Intelligence in Electromyography Analysis-A Perspective on Current Applications and Future Challenges*, 3–36.

[91] **Allard, U.C.**, **Nougarou, F.**, **Fall, C.L.**, **Giguère, P.**, **Gosselin, C.**, **Laviolette, F. and Gosselin, B.** (2016). A convolutional neural network for robotic arm guidance using sEMG based frequency-features, *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp.2464–2470.

[92] **van der Maaten, L. and Hinton, G.** (2008). Viualizing data using t-SNE, *Journal of Machine Learning Research*, *9*, 2579–2605.

[93] **Popovici, I.** (2020). Experimental Results on the Accuracy of the Myo Armband for Short-Range Pointing Tasks, *2020 International Conference on Development and Application Systems (DAS)*, IEEE, pp.185–188.

[94] **Popovici, I. and Vatavu, R.D.** (2018). Perceived Usability, Desirability, and Workload of Mid-Air Gesture Control for Smart TVs., *RoCHI*, pp.91–98.

**CURRICULUM VITAE**

**Name Surname: Mustafa Seddiqi**

**EDUCATION:**

- **B.Sc.:** 2017, Middle East Technical University, Applied Science, Electrical and Electronics Engineering
- **M.Sc.:** 2021, Istanbul Technical University, Applied Science, Computer Engineering

**PROFESSIONAL EXPERIENCE AND REWARDS:**

- 01/2020 - present Research & Development Engineer at ELKON Concrete Batching Plants, Tekirdag, Turkey
- 04/2018 – 09/2018 Research & Development Engineer at 3D3 Teknoloji METU-Teknokent, Ankara, Turkey
- 2010 Ranked in the top 1% among more than 120000 students entering Kankoor (Afghanistan University Entrance Exam)

**PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:**

- M. Seddiqi, H. Kivrak and H. Kose, "Recognition of Turkish Sign Language (TID) Using sEMG Sensor," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey.