Project Report On

MyoAuth: Biometric Authentication using Surface Electromyography

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

By

Nikhil Zachariah (U2003153)
Niya Bimal (U2003156)
Noel Joseph Paul (U2003158)
Sanjana Nair (U2003187)

Under the guidance of

Ms. Dincy Paul

Department of Computer Science and Engineering
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)
Rajagiri Valley, Kakkanad, Kochi, 682039

April 2024

CERTIFICATE

This is to certify that the project report entitled ***"MyoAuth"*** *is a bonafide record of the work done by* ***Nikhil Zachariah (U2003153), Niya Bimal (U2003156), Noel Joseph Paul (U2003158), Sanjana Nair (U2003187)****, submitted to the Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in Computer Science and Engineering during the academic year 2023-2024.*

Ms. Dincy Paul	Ms. Anita John
Project Guide	Project Coordinator
Assistant Professor	Assistant Professor
Dept. of CSE	Dept. of CSE
RSET	RSET

Dr. Preetha K.G.
Head of Department
Professor
Dept. of CSE
RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Prof.(Dr.) P. S. Sreejith**, Principal of RSET, and **Dr. Preetha K.G.**, Head of the Department of Computer Science and Engineering, for providing us with the opportunity to undertake our project, **MyoAuth**.

We are highly indebted to our project coordinators, **Ms. Anita John**, Assistant Professor, Department of Computer Science and Engineering and **Mr. Sajanraj T.D.**, Assistant Professor, Department of Computer Science and Engineering, for their valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our project guide **Ms. Dincy Paul**, Assistant Professor, Department of Computer Science and Engineering, for her patience and all the priceless advice and wisdom she has shared with us.

Last but not least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

Nikhil Zachariah
Niya Bimal
Noel Joseph Paul
Sanjana Nair

Abstract

As technology advances, privacy concerns, including security breaches in biometric authentication systems, have increased. Current systems, relying on fingerprint, iris, or face scans, are susceptible to spoofing attacks using fake biometric data. This includes 3D models of fingerprints and forged images or videos for facial recognition. Cost concerns for more secure biometrics like iris scans also persist.

Surface Electromyography or sEMG biometric authentication is a cost-efficient and non-invasive technique and is considered much more robust and difficult to forge compared to other biometrics. This is because, unlike static and unique biometrics like fingerprint and facial recognition, sEMG is dynamic and real-time. Hence replicating a specific muscle activation pattern is very challenging. Another unique feature of sEMG is that it can also provide liveness detection, ie. involuntary or forced muscle movements will not produce the same sEMG signals required to authenticate the user and most likely would produce patterns corresponding to noise. Hence forceful or involuntary authentication attempts are also not permitted.

To demonstrate the idea of this research we have prepared a minimal expense prototype model. The prototype contains modules for signal acquisition, signal processing, segmentation, feature extraction and authentication of users using a KNN (k-nearest neighbors algorithm) integrated ML model. The prototype features a UI that allows user to register their details following which they are prompted to register their sEMG for a specific gesture five times using the signal acquisition hardware module. The system proceeds to perform advanced signal processing techniques. Unique features corresponding to the acquired signals are extracted and combined to form a feature vector for each user. These feature vectors are trained by the ML model that learns the feature vector and its associated username. The trained model can then be used to identify the user using a new set of EMG signals.

Contents

List of Abbreviations

EMG - Electomyography

sEMG - Surface Electomyography

TD - Time Domain

FDT - Frequency Division Technique

AR - Autoregressive Coefficients

KNN - K-nearest neighbours

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The evolution of biometric technologies has revolutionized the landscape of personal identification, addressing security and convenience concerns across various domains. Traditional methods, such as passwords and PINs, have progressively given way to more sophisticated biometric approaches like fingerprint recognition and facial identification. While these technologies have demonstrated significant advancements and widespread adoption, they are not without limitations.

Fingerprint recognition, although widely used, can be hindered by issues such as skin conditions, environmental factors, and the potential for unauthorized access through forged prints. Similarly, facial recognition, though increasingly prevalent, faces challenges related to accuracy, susceptibility to environmental conditions, and privacy concerns regarding mass surveillance. Recognizing these challenges has driven the search for innovative alternatives, and among them, Electromyography (EMG) signals emerge as a promising candidate.

1.2 Problem Definition

The existing landscape of bio-metric identification technologies, including fingerprint and facial recognition systems, faces inherent limitations such as susceptibility to environmental conditions, privacy concerns, and the potential for unauthorized access. These challenges necessitate the exploration of alternative, more secure, and adaptable bio-metric solutions.

The problem at hand is to develop an innovative identification system that overcomes the drawbacks of existing technologies. Specifically, the project aims to leverage Electromyography (EMG) signals, capturing the electrical activity of muscles, to create a biometric system that excels in liveness detection, individual recognition, and adaptability. The overarching problem is to enhance the reliability, security, and user-friendliness of biometric identification, addressing the shortcomings of current methods through the unique advantages offered by EMG signals.

1.3 Scope and Motivation

1.3.1 Scope

The scope of this project is expansive, covering a multifaceted exploration of EMG signal applications. It includes an in-depth investigation into precise electrode placement techniques, an exploration of state-of-the-art signal processing methodologies, and the subsequent development of a system for personalized identification models. While the primary focus is on person identification, the implications extend to potential applications in healthcare and security systems. The project's scope is not only ambitious but also aligned with the broader technological landscape, positioning itself as a pioneering effort in the integration of biometrics and electromyography.

1.3.2 Motivation

The motivation behind this project arises from a critical evaluation of the drawbacks associated with existing biometric technologies. While fingerprint and facial recognition have proven effective in numerous applications, their limitations prompt the exploration of alternative methods. The motivation is rooted in the quest for a biometric solution that overcomes the challenges of current technologies, providing a more secure, adaptable, and user-friendly approach to person identification.

The inherent drawbacks of existing biometric methods, including susceptibility to environmental factors, privacy concerns, and the potential for unauthorized access, fuel the drive to explore Electromyography (EMG) signals as a viable alternative. The advantages of EMG signals extend beyond conventional biometrics. Liveness detection, a crucial aspect of biometric systems, is inherently addressed by EMG signals as they capture real-time muscle activity. This capability adds an extra layer of security by ensuring that the presented signal is a dynamic, live response, mitigating the risk of spoofing or unauthorized access. Furthermore, EMG signals offer a nuanced level of personalization, allowing for the recognition of individuals based on unique muscle activity patterns.

This motivates us to delve into the untapped potential of EMG signals, with the anticipation that it could address the limitations of current biometric technologies and pave the way for a more robust and versatile identification system.

1.4 Objectives

The objectives of this research project are two fold.

- Firstly, it aims to conduct an extensive investigation into EMG signal acquisition techniques, emphasizing the refinement of precise electrode placement and the integration of advanced signal processing methods. This involves a nuanced exploration of various methodologies to capture, process, and interpret EMG signals accurately.
- Secondly, the project seeks to develop a robust identification system capable of uniquely identifying individuals based on the distinct patterns within their muscle activity. These objectives collectively contribute to advancing the field of biometric technology, specifically in the domain of secure and non-intrusive person identification.

1.5 Purpose and Need

The purpose of this project is to address the pressing need for a more reliable, user-friendly, and secure identification system. Traditional methods often fall short in providing a foolproof solution, and the need for innovation in this space is critical. The purpose of this research is not just academic; it is driven by a practical necessity for identification systems that transcend the limitations of current technologies. This research serves the purpose of fulfilling this need by exploring the uncharted territory of EMG signals, aiming to harness their unique properties for effective person identification.

1.6 Challenges

Noise and Interference: EMG signals can be affected by various types of noise, including environmental interference, electrode motion artifacts, and electrical noise. Ensuring signal quality and filtering out unwanted noise is a significant challenge.

Dynamic Nature of Muscle Activity: Muscles are dynamic and can undergo changes due to fatigue, injury, or other factors. Adapting the model to accommodate these dynamic changes is a challenge.

Subject-Specific Adaptation: Some individuals may have unique muscle activation patterns that require personalized models. Designing a system that can adapt to individual differences is challenging.

Real-time Processing: Depending on the application, real-time processing of EMG signals may be crucial. Achieving low-latency processing while maintaining accuracy is a challenge, especially in resource-constrained environments.

Privacy and Ethical Concerns: Collecting and utilizing biometric data, such as EMG signals, raises privacy and ethical concerns. Ensuring user consent, data security, and compliance with regulations is essential.

Limited Training Data: Obtaining a sufficiently large and diverse dataset for training the KNN classifier may be challenging, especially for specific populations or activities.

Cross-Subject Generalization: Ensuring that the model generalizes well to new subjects who were not part of the training set is a common challenge in biometric identification systems.

User Acceptance and Comfort: Users may find wearing EMG electrodes uncomfortable or invasive, leading to potential usability issues and a decrease in user acceptance.

1.7 Assumptions

Consistency of EMG Signals: The assumption is that EMG signals for the same person remain consistent over time. Changes in muscle condition, fatigue, or other physiological factors might affect the consistency.

Individual Variation: The assumption is that individuals have distinct EMG signal patterns. This assumption may be affected by factors such as age, gender, and individual variations in muscle anatomy.

Training Set Representativeness: The training dataset used for the k-NN classifier should be representative of the population. It should cover a diverse range of individuals, including various activities and muscle movements to ensure a robust model.

Stability of Electrode Placement: The accuracy of the model assumes stable and accurate placement of electrodes on the muscles. Any variation in electrode placement can introduce noise and affect the reliability of the system.

1.8 Societal / Industrial Relevance

In an era where technological advancements shape societal structures, the social and industrial relevance of this research cannot be overstated. The integration of biometrics and electromyography holds immense potential for societal transformation. In the realm of security, this research can contribute to the development of systems that safeguard sensitive information and infrastructures. Moreover, the implications extend to the healthcare sector, where the technology could be applied in prosthetics control, assistive devices, and rehabilitation. The social impact of this project is not limited to the technology sector. it encompasses improved user experience, enhanced security, and potential innovations in healthcare.

1.9 Organization of the Report

The organization of this report is designed to provide a coherent and detailed account of the project. Following this comprehensive introduction outlined in Chapter 1, we have a detailed literature survey on the various studies and findings we received from existing research works in Chapter 2. It also outlines the gaps identified in the existing biometric and non-biometric authentication systems. Chapter 3 specifies the hardware and software requirements of this project, functional requirements and also mentions the costs that have incurred during the course of the project. System architecture and phases of working are discussed in detail in Chapter 4. The timeline and workflow of this project is also outlined using a Gantt chart in this section. Chapter 5 dives into the methodology implemented in this project. It also outlines the various datasets and dicusses about the implementation strategies. Chapter 6 gives in detail the results of the project mentioning the accuracy scores and screenshots of the project interface. It also discusses comparison with other models and their accuracy. Chapter 7 gives a detailed conclusion, summarizing the project's findings and achievements.

Chapter 2

Literature Survey

2.1 A Study of Personal Recognition Method Based on EMG Signal

2.1.1 Introduction

Personal recognition methods using traditional identification methods such as PIN, ID, or signature are not reliable due to the risk of leakage, theft, imitation, and forgery.

Biometric technology, such as face, fingerprint, iris, and voice recognition, has been widely used to eliminate the risks of traditional methods, but they also have weaknesses such as counterfeiting and falsification. The study proposes EMG-based personal identification methods using DWT and ExtraTreesClassifier, and later CWT and CNN to overcome the limitations of re-training the model. An EMG-based verification technique using CWT and siamese networks is also proposed.

2.1.2 Datasets

The study collected surface EMG signals from 21 subjects using the Myo armband placed on the right forearm.The EMG signals were collected under the hand-open gesture.Two different methods were proposed for EMG-based personal identification, and experiments were conducted using these methods with the 21 subjects.The identification accuracy of the methods was evaluated using these datasets, achieving accuracies of 99.206% and 99.203% respectively. The study also used a validation set to test the performance of different mother wavelet functions for feature extraction. Several widely used mother wavelet functions were tested on the validation set to find the appropriate one. The study applied maxpooling on the input before performing a convolution to reduce the computational load of the ConvNet during training and inference. The kernel size of the convolutional layer varied for different layers, with the first three layers having a size of 3x3 and the last layer having a size of 2x6. The study used the DWT method for data preprocessing and feature extraction, which involved decomposing the EMG signals into frequency sub-bands. The CWT method was used for data preprocessing in the EMG-based personal identification method, which involved time-frequency analysis using a set of functions generated by the mother wavelet.

2.1.3 Method

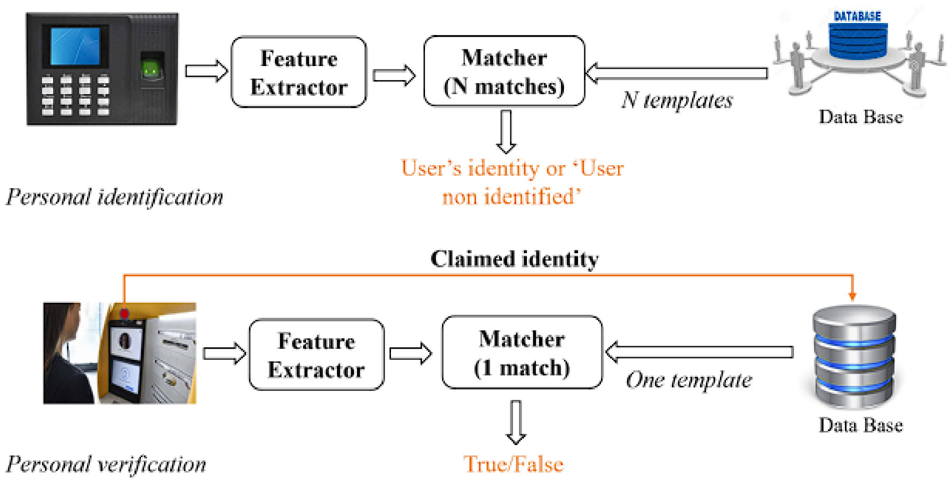


Figure 2.1: Block diagrams of personal identification mode and personal verification mode

The proposed method for personal identification is based on electromyography (EMG) signals and involves two different approaches: one based on Discrete Wavelet Transform (DWT) and ExtraTreesClassifier, and the other based on Continuous Wavelet Transform (CWT) and Convolutional Neural Networks (CNN) .The EMG signal is captured from the arm of the subjects using a Myo armband .The collected data is pre-processed using DWT to extract features .The ExtraTreesClassifier algorithm is then used to classify the subjects For model update when new data is added, a transfer learning algorithm is adopted based on the identification method using CWT and CNN . An EMG-based personal verification method is proposed using CWT and siamese networks.

2.1.4 Evaluation

The proposed EMG-based personal identification methods using DWT and ExtraTreesClassifier, as well as CWT and CNN, achieved high identification accuracies of 99.206% and 99.203% respectively .The EMG-based personal verification method using CWT and siamese networks achieved a verification accuracy of 99.285% .The evaluation was conducted on a dataset of 21 subjects .The study did not provide information on the performance of the methods on larger datasets or their scalability .The limitations of the EMG-based recognition system, such as its suitability for individuals with neurodegenerative disorders, were mentioned but not evaluated in the study.

2.2 EmgAuth: Unlocking Smartphones With EMG Signals

Deep neural networks, renowned for their prowess in tasks like image classification, speech recognition, and natural language processing, excel at autonomously deriving meaningful features from extensive datasets, eliminating the need for manual feature engineering.

Siamese networks, a concept pioneered by Bromley et al., leverage twin sub-networks to extract features and gauge the disparity between their outputs. This unique architecture renders them adept at handling smaller datasets and specialized tasks such as signature verification, video-based object detection, human identification through gait recognition, one-shot image recognition, and fine-grained relation extraction. Demonstrating remarkable efficacy in feature extraction with limited samples, Siamese networks have become an invaluable technique across diverse domains.

2.2.1 Architecture

EmgAuth presents a comprehensive system comprising components integrated into a Myo armband and an Android smartphone. The Myo armband, equipped with eight sensors and operating at a sample rate of 200 Hz, gathers EMG signals. Two distinct Android applications facilitate data collection, labeling, and the simulation of an unlocking process. To bring the deep learning model into action, TensorFlow Mobile is utilized, with the model itself trained on a GPU server. The system's architecture encompasses offline model training, data segmentation, and data augmentation techniques. Specifically, a convolutional Siamese neural network is trained by utilizing pairs of valid EMG signals. During online authentication, four distinct sets of motions are enrolled, and incoming EMG signals are compared against the previously stored signals. The Siamese neural network computes the distance between input EMG signal pairs, enabling the system togrant or deny authorization to the user based on this analysis

2.2.2 Dataset

The EmgAuth system relies on an Electromyography (EMG) signal dataset collected from multiple users wearing Myo armbands. These signals, recorded from various positions on users' forearms, capture muscle activity during hand movements, forming the foundation of the dataset. To enrich this dataset, common image classification data augmentation techniques like flipping, cropping, and scaling are employed. These augmentations generate diverse variations of the original EMG data while preserving labels and user identities. This augmented dataset significantly enhances the capacity of the deep neural network to discern additional features and make consistent decisions, irrespective of how the Myo armband is positioned or rotated by the user. The amalgamation of the original and augmented datasets empowers the EmgAuth system to accurately identify users based on their EMG signals, demonstrating robustness even in scenarios involving arm movement or rotation

2.2.3 Evaluation

The evaluation of the EmgAuth system encompasses key metrics like accuracy, true acceptance rate (TAR), false acceptance rate (FAR), false rejection rate (FRR), and equal error rate (EER). Through 5-fold cross-validation, the system achieves an overall average accuracy of 92.06%, coupled with a 91.81% TAR, 7.43% FAR, and 8.49% FRR. However, a noticeable decline in accuracy to 87.50% is observed within the third group, significantly impacting the system's overall performance. This variability in metrics during cross-validation is primarily attributed to two factors: time drifting, leading to mislabeling due to users not adhering to the sampling time, and the collection of unsuitable data arising from participants executing movements in an unnatural manner. Notably, the paper lacks detailed insights into how different hyperparameters influence the system's performance, leaving this aspect unexplored in the analysis. In determining the classification threshold for the EmgAuth system, the paper employs the Detection Error Trade-off (DET) curve analysis. The threshold corresponding to the Equal Error Rate (EER) point, specifically at 0.55, is selected as the final threshold for the system. This choice aligns with the EER point, where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). The rationale behind opting for the EER threshold is rooted in the system's objective to strike a balance between security and efficiency. While a reduction in FAR could enhance security, it might also impact the efficiency of the unlocking process. Hence, the emphasis is placed on improving efficiency and convenience for users. By setting the EER threshold at 0.55, the system achieves a balance between security and convenience, allowing swift unlocking of the smartphone as soon as the user picks it up.

2.3 EMG Biometric Systems Based on Different Wrist Hand Movements

2.3.1 Introduction

EMG-based biometric authentication can be applied in robot control, allowing only authorized individuals to operate robots . Prosthetic devices can benefit from EMG authentication, ensuring secure and personalized control for the intended user . EMG authentication can enhance the security and user experience in virtual reality environments, enabling secure access and interaction with virtual worlds . Wearable devices, such as smartwatches or fitness trackers, can incorporate EMG authentication for secure access and personalized control features.

2.3.2 Datasets

The paper utilizes three EMG datasets with similar EMG sensing in different sessions for the development of an extensive biometric system The first dataset consists of recordings from the first week, with 20 sessions and 160 windows per movement, which are used for training The second dataset, by Raurale et al., includes EMG recordings from ten subjects, with twenty sessions recorded from each subject. Each session consists of continuous EMG recordings of nine movements in a fixed order . The third dataset, by Angeles et al., includes EMG recordings of fifty healthy subjects performing ten different movements using the same armband. Each subject performed each movement five times, resulting in a total of sixteen seconds of EMG recording . These datasets provide a diverse range of

EMG recordings from different subjects and movements, allowing for the development of a robust biometric system.

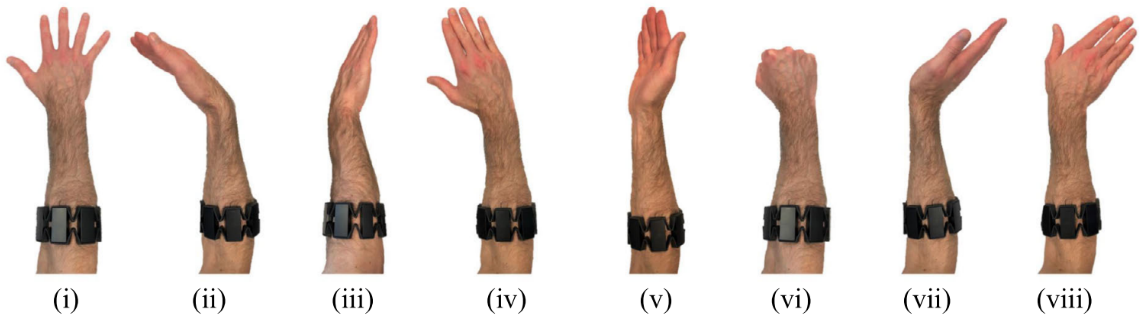


Figure 2.2: Considered wrist hand movements

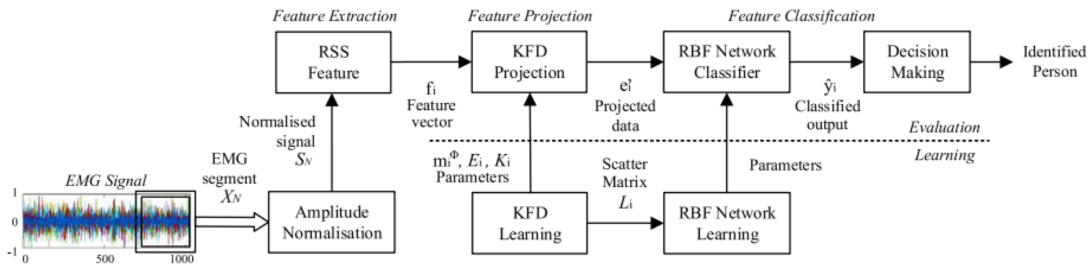


Figure 2.3: Proposed Biometric Identification System

2.3.3 Method and Evaluation

EMG signals can be used for biometric identification by normalizing the amplitude and extracting features. Root sum square (RSS) is used to represent muscle contraction level and derive a unique value for each person. Non-linear projection schemes like KPCA and KFD are used for feature projection and classification. The RBF neural network with KFD projection analysis shows the highest AUC and is selected for the system pipeline. The proposed EMG biometric identification system architecture is summarized. The system evaluation phase involves training the KFD model and evaluating accuracy on trained RBF network parameters. On-going identification performance is analyzed for different re-identification intervals. The proposed system achieves identification accuracy of over 91% with an EER value of around 17%. The system can be implemented in real-time on an ARM Cortex-A53 platform with processing times of 1.06 ms for verification and 1.61 ms for identification. The proposed biometric system uses forearm EMG for accurate verification and identification using time-domain pattern recognition. It acquires EMG from randomly-placed sensors and employs time-domain features followed by LDA-based 8 projection and MLP classification. The one-time verification system shows over 92% accuracy with 27.79% EER, while the progressive on-going verification system achieves up to 93% re-verification accuracy. For identification, 256-sample EMG segments are normalized and projected using KFD and classified by RBF-NN. The identification system achieves 91% accuracy with 17.28% EER in one-time verification and 92% accuracy with 16.42% EER in progressive on-going identification. The proposed authentication system requires 1.06 ms to execute per 256-sample EMG window on an ARM Cortex-A53 platform. The system can be realized in real-time on battery-operated hardware platforms without repeating the same wrist-hand pose.

2.4 Challenges of Synthetic EMG Based on Adversarial Style Transfer

The paper defines two evaluation metrics, namely hit rate and confusion rate, to quantifiably evaluate the success rate of using synthetic EMG signals to attack identification models and the chance of models being confused by the synthetic signals.

The paper also discusses the training approach of the generator network, which can involve using a single network to generate signals of different categories or setting up multiple networks, with each network responsible for generating data of one category. The latter approach typically yields better generation results.

Hit rate is defined as the success rate of using synthetic EMG signals to attack identification models, and confusion rate as the chance of identification models being confused by the synthetic EMG signals. The hit rate of the attack methods on three different identification models (GengNet, EMGNet, and VGG16Net) were 89.34, 87.94, and 97.24, respectively. The confusion rate of the attack methods on three different identification models (GengNet, EMGNet, and VGG16Net) were 99.06, 99.17, and 100, respectively.

2.4.1 Methods

This paper explores two EMG-based personal identification methods:

1. A method that uses Discrete Wavelet Transform (DWT) and ExtraTreesClassifier.
2. A method that incorporates Continuous Wavelet Transform (CWT) and Convolutional

Neural Networks (CNN).

Both methods achieved remarkable identification accuracies of approximately 99.206 percent and 99.203 percent respectively. The authors also introduce a transfer learning algorithm to address the challenge of model updates, building upon the identification method employing CWT and CNN.

2.4.2 Applications

- Adversarial Testing: Synthetic EMG signals can be utilized in ethical hacking scenarios or security testing environments. They can serve as a tool to evaluate the resilience of EMG-based biometric systems against adversarial attacks, helping researchers and developers identify weaknesses and improve system security.

- Development of Attack Strategies: Synthetic EMG signals enable researchers to experiment with various attack strategies. By creating and testing these signals, attackers can develop more sophisticated methods to deceive EMG-based authentication systems, potentially exploiting vulnerabilities for unauthorized access.

- Training and Defense Enhancement: Synthetic EMG signals can aid in training machine learning models to recognize and defend against adversarial attacks. They can be used as part of the training dataset to enhance the system's ability to distinguish between genuine and synthetic signals, thereby improving its robustness against future attacks.

2.5 Performance Optimization of sEMG based Biometric System

2.5.1 Abstract

This literature survey explores a recent paper that investigates three distinct feature sets for surface electromyography (sEMG)-based biometric authentication. The focal points are the Frequency Division Technique (FDT), Time-domain (TD), and Autoregressive (AR) features. The study presents a comprehensive analysis of each feature set's performance across multiple channels, emphasizing their potential implications for real-world applications.

2.5.2 Hudgin's Time-Domain (TD) Method

The TD feature set focuses on time-domain characteristics of the sEMG signal. This set is characterized by parameters such as mean absolute value, waveform length, and other time-based features. The literature survey aims to elucidate the specific TD features considered in the comparative analysis and their relevance to sEMG-based authentication.

$$RMS = \sqrt{\tfrac{i}{N} \sum_{i=1}^{N} x_i^2}$$

Root Mean Square : measure of amplitude
i - index of the sample, N - no. of samples, x - signal.

$$WL = \sum_{i=1}^{N-1} |x_{i-1} - x_i|$$

Waveform Length: measure of total length of a signal.
It is calculated by summing absolute differences between adjacent samples.

$$SSO = \sum_{i=2}^{N-1} f(x_i - x_{i-1})(x_i - x_{i+1})$$

Sum of Squared Order: measures signal smoothness.
It's calculated by summing the product of differences between adjacent samples and the difference between the sample and its two neighbors.

Figure 2.4: Time-domain features that can be extracted for TD feature set

2.5.3 Frequency Division Technique (FDT)

In contrast to TD, the FDT feature set is defined by an equation that involves the summation of magnitudes of the Fourier Transform of the sEMG signal at specific frequencies denoted by f(i,j). For each channel i, the frequencies range from 1 to N(i). The resulting sum undergoes processing through a function F to obtain the FDT feature for that channel.

2.5.4 Autoregressive (AR) Feature Set

The AR feature set involves modeling the sEMG signal using autoregressive parameters. This set captures the temporal dependencies within the signal, providing a different perspective on muscle activation patterns. The survey investigates the specific AR features

$$FDT_i = F\left[\sum_{j=1}^{N_i} |X(f_{i,j})|\right], \qquad i = 1, 2 \ldots L$$

Defines FDT feature set for the **i-th channel** (having several frequencies).

By summing the absolute values of the Fourier Transform (F) of the signal **at each frequency (f)** of each channel

j ranges from 1 to the total number of **frequencies** N(i) for the i-th channel.

The no. of frequencies used for each channel (N) varies and is denoted by Ni.

Figure 2.5: Division of signals into distinct frequency sub-bands in FDT method

considered in the paper and their impact on the overall performance of the authentication system.

$$x_i = \sum_{p=1}^{P} a_p x_{i-p} + w_i$$

To model time-series data, where the current value of the time series is dependent on its past values.

xi & x(i–p) - is the value of the time series at time *i and (i-p)*
ap - coefficient of the *p*-th lagged value of the time series
w(i) is the error term at time *i*.

The coefficients ***ap*** represent the **strength of the relationship** between the current value and the past values of the time series.

Figure 2.6: AR model - a prediction model

2.5.5 Evaluation

The results from the paper indicate that the TD feature set consistently outperformed both FDT and AR across all channel numbers. This finding raises questions about the effectiveness of frequency-based features in comparison to time-domain features for sEMG-based biometric authentication. The paper also discusses potential avenues for future research in refining feature sets and enhancing the robustness of sEMG authentication systems.

2.6 Summary and Gaps Identified

2.6.1 Summary of Literature Survey

Paper	Description	Advantage	Disadvantage
[1]	EMG-based personal identification method using Discrete Wavelet Transform (DWT) and ExtraTreesClassifier algorithm.	Achieves high identification accuracy of 99.206% and is feasible for personal identification based on EMG signals.	Requires re-training of the whole classifier when new data is added, which consumes a significant amount of time and computation resources.
[2]	The paper discusses related work on biometric authentication methods, EMGbased applications, and the Siamese network, and provides details on the EmgAuth system architecture	Learn the difference between a pair of inputs, which makes it well-suited for tasks that involve comparing EMG signals from different users.	Limited Comparative Analysis: May overlook potential alternative methods. Empirical Validation: Requires further real world testing for practical application.
[3]	Acquire EMG signals from the user's wrist-hand movements, extracts time-domain features from the signals, and uses a Radial Basis Function Neural Network (RBF-NN) classifier to classify the features into different classes.	Simple to train and require fewer training samples compared to other types of neural networks	RBF-NNs can be computationally expensive, especially when dealing with large datasets or high dimensional feature spaces
[4]	Introduces an attack method based on generative strategies to generate synthetic EMG hand gesture signals, highlighting the potential risks of synthetic biological signals in identification systems.	Demonstrate the feasibility of the proposed method and highlight the vulnerability of deep EMG classifiers	The paper does not discuss the potential limitations or challenges associated with implementing the GAN
[5]	Performance comparison of different feature extraction methods: Hudgin's time domain, Frequency Division Technique (FDT), and Autoregressive Coefficients (AR)	TD feature set offers superior authentication sensitivity and robustness. Four-channel configuration balances performance and complexity effectively.	Limited Comparative Analysis: May overlook potential alternative methods. Empirical Validation: Requires further real-world testing for practical application.

Table 2.6.1: Summary of Literature Survey

2.6.2 Gaps Identified with existing systems

Biometric Method	Disadvantages/Challenges
Face	- Vulnerable to spoofing attacks using high-resolution photos or videos. Accuracy can be affected by lighting and environmental conditions. Privacy concerns due to biometric data storage.
Iris	- High-cost implementation due to specialized hardware requirements. Potential for spoofing using high-quality images or contact lenses. Limited availability of iris scanners.
Fingerprint	- Vulnerable to spoofing using fake fingerprints or lifted prints. Dirty or damaged fingers may lead to authentication failures. Limited accuracy in extreme weather conditions.

Table 2.6.2: Gaps identified with existing biometric systems

Non-Biometric Method	Disadvantages/Challenges
PIN	- Susceptible to brute force attacks if the PIN is too short or easily guessable. Risk of shoulder surfing or observation-based attacks. Users may forget or lose their PIN.
Password	- Prone to dictionary attacks if passwords are weak or common. Users often reuse passwords across multiple accounts, increasing security risks. Password phishing and social engineering attacks.
Pattern	- Vulnerable to shoulder surfing or smudge attacks on touchscreen devices. Lack of complexity compared to passwords, reducing security. Users may forget their pattern or find it cumbersome.
Token	- Risk of token theft or loss, compromising authentication security. Token hardware may be expensive to maintain and replace. Potential for unauthorized access if tokens are shared or stolen.
Smart Card	- Vulnerable to physical theft or cloning of smart cards. Requires specialized card readers and infrastructure for authentication. Risk of card damage or malfunction leading to authentication failures.

Table 2.6.3: Gaps identified with existing nonbiometric systems

Chapter 3

Requirements

3.1 Hardware Requirements

- Amplifiers: BioAmp is used to amplify weak EMG signals for proper processing and to boost the signal-to-noise ratio.
- Arduino board
- Computing Resources: 200 GB disk space and 8 GB RAM minimum.
- EMG Sensors: EMG acquisition band or electrodes, electrode gel, and wires for connections. For this project we used the EMG signal acquisition band, BioAmp from UpsideDown Labs.

Features & Specifications:

Length: 13 inches
Stretchability: 2X (Upto 26 inches)
Usability: Reusable as it comes with washable fabric
Interface: Snap electrodes
Compatible Hardware: BioAmp Cable used with BioAmp Hardware (Muscle BioAmp Candy)
BioPotentials: EMG
No. of channels: 1
Wearable: Yes

3.2 Software Requirements

- Signal Processing Software: Software tools for preprocessing and cleaning the raw EMG signals. Involves using signal processing libraries or tools like MATLAB or Python with libraries such as SciPy.
- Machine Learning Libraries: Python programming language and its machine learning libraries (e.g., scikit-learn) for implementing and training the k-NN classifier.
- Integrated Development Environment (IDE): An IDE such as Jupyter Notebooks, Google Colab or any preferred Python development environment for coding and testing the algorithms.

3.3 Budget

Hardware:
Arduino UNO : Rs. 650
EMG kit (Muscle BioAmp Candy, EMG band, Cable wires x3) : Rs. 800
Electrode patches (100 nos.) : Rs. 400
Electrode gel : Rs. 150

Documentation and Reports:
Printing costs: Rs. 5000

Miscellaneous:
Travel for project-related activities: Rs. 2000
Contingency fund for unforeseen expenses: Rs.1000

Approx. Total: Rs. 10,000

Chapter 4

System Architecture

4.1 System Overview

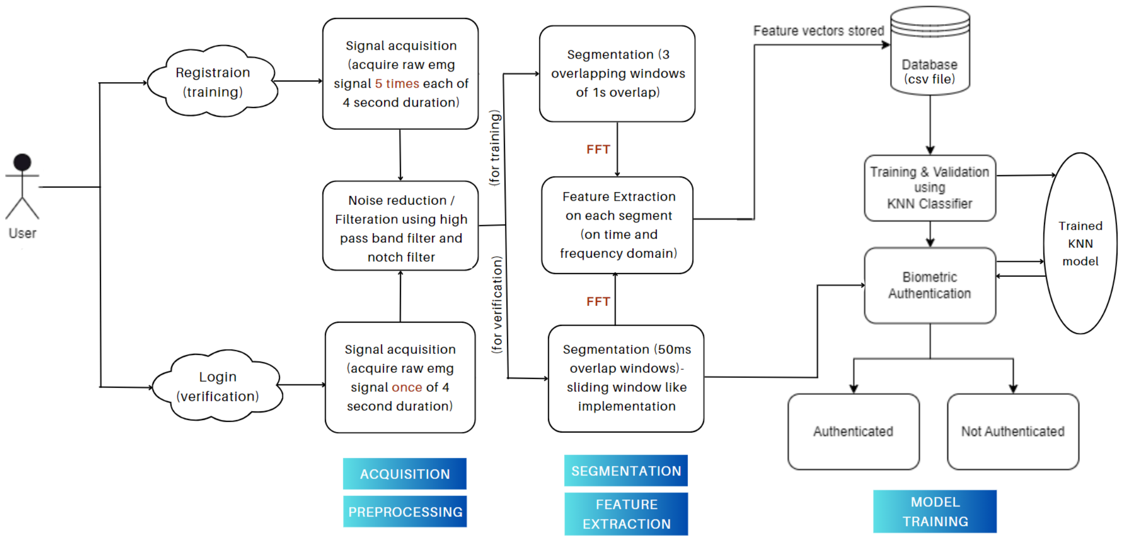


Figure 4.1: Architecture Diagram

The architecture of this biometric authentication system using surface EMG signals begins with user interaction, where users can either register for training or log in for verification.

During registration, the system acquires the raw EMG signal five times, each for a duration of four seconds, capturing unique muscle activity associated with specific gestures. For login purposes, the system acquires the raw EMG signal once, also for a duration of four seconds. After signal acquisition, the system applies noise reduction and filtration techniques, such as using a low-pass band filter, to eliminate unwanted signals and artifacts from the EMG data. Following this, the segmented approach varies for registration and login processes. During registration, the system segments the signal into three overlapping windows with a one-second overlap between each window, while for login, it implements segmentation with 50ms overlap windows, akin to a sliding window technique. Subsequently, feature extraction is performed on each segment, extracting features from both the time and frequency domains of the EMG signal. A Fast Fourier Transform is used to convert from time domain to frequency domain. These extracted features are then stored in a CSV file for further analysis and processing.

The final stage involves biometric authentication where the feature vectors stored in the CSV file are utilized to train a K-Nearest Neighbors (KNN) classifier. This classifier learns patterns and characteristics from the extracted features enabling it to authenticate or reject a user based on their EMG signal patterns, thereby ensuring secure and reliable biometric authentication.

This comprehensive process ensures the robustness and reliability of the EMG-based biometric authentication system, effectively securing access based on distinct features within the EMG signals.

4.2 Architectural Design

4.2.1 Sequence Diagram

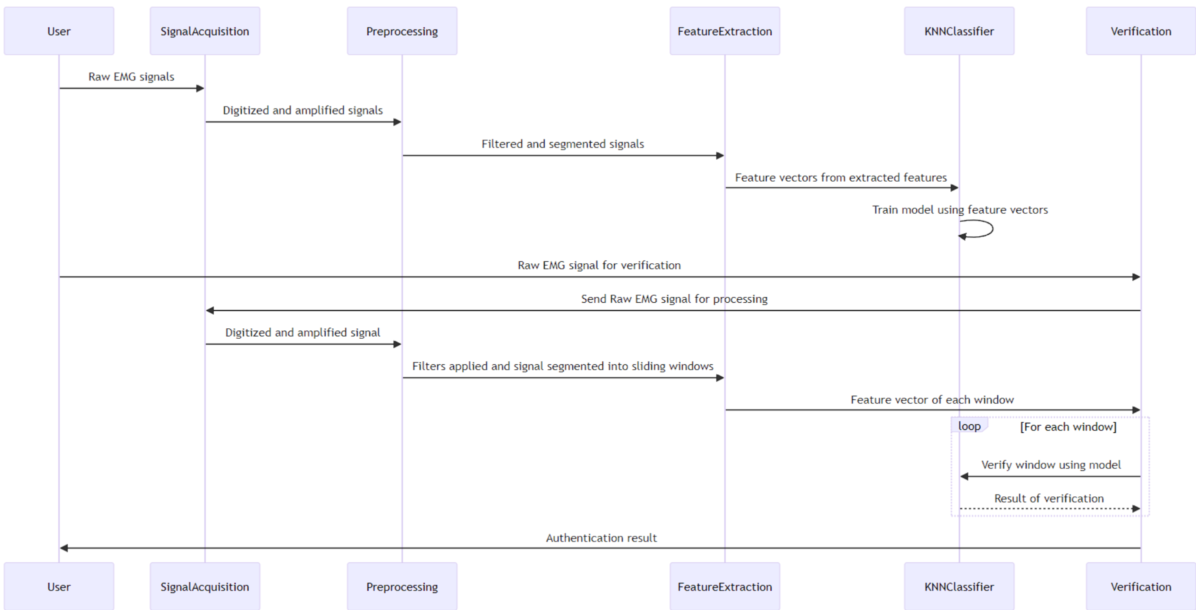


Figure 4.2: Sequence Diagram

- **Registration Phase:**

  In the signal acquisition stage,raw EMG signals are acquired from the user wearing the band. The acquired signals are digitized and amplified. It is then sent to the signal preprocessing module. Here a high pass filter of 5 Hz and a notch filter of 60 Hz is applied. These allow only a range of frequencies to pass through and eliminate noise. The data is also segmented to analyze certain sections of the signal and isolate relevant muscle activity from noise. Twenty one features are extracted in both time and frequency domain.In EMG signal processing, feature extraction condenses raw data into relevant characteristics. These features enable analysis for tasks such as muscle activity classification. Feature vectors are made to make it easier for machine learning models to analyze the values.The labeled feature vectors are fed into KNN model for training.

- **Login Phase:**

  Raw EMG Signal is recorded by the user.When the verification model receives it, it sends the data for processing. The signal is digitized to convert the continuous analog signals produced by muscle activity into discrete digital values, enabling efficient processing and storage. It is then amplified making them easier to detect and analyze, especially in situations where the signal strength is low or noise levels are high.After this, filters are applied and segmented into sliding windows for temporal analysis, feature extraction, and noise reduction by dividing the signal into smaller, overlapping windows. Feature vectors are extracted from each window and verification is done by passing each window through the trained KNN model. The authentication result is then sent to the user.

4.3 Work Schedule - Gantt Chart

Figure 4.3: Gantt Chart

Chapter 5

System Implementation

This chapter describes in detail the methodology used in this project and how it was implemented. It details the working of the prototype developed as part of this project.

5.1 Datasets Identified

- **Raw CSV file:** For the creation of dataset, users are asked to register a gesture of 4 second duration, 5 times. The recorded raw EMG data is saved to a CSV file containing the fields, (Username, Timestamp, EMG values). Therefore 5 entries are present per user. During login phase signal is acquired only once for verification thus only 5 entries will be present.

- **Filtered CSV file:** High pass band filter and notch filter are applied on the raw EMG values to get the filtered EMG values.

- **Segmented CSV during registration:** The filtered CSV file undergoes segmentation such that 1 EMG singal is segmented to 3 overlapping segments. The overlapping segments are saved in a CSV file containing the information, (Username, Window number, Start time, End time, EMG values). Therefore 1 filtered EMG signal gets divided to 3. In registration phase, 5 registered EMG signals of a user gets divided to 15 segments/entries.

- **Feature vector:** 21 features are extracted from a segment and appended to a feature vector. Similarly feature vectors are created for every segment.

5.2 Proposed Methodology

5.2.1 Signal Acquisition

The initial step of our project involves acquiring electromyography (EMG) signals from users using a specialized single-channel EMG recording wearable band. This band comprises three sensors, with one sensor strategically positioned on the ulnar nerve to capture unique EMG signals associated with muscle movements. This setup not only enables precise authentication but also ensures the reliability of signal acquisition. A simple electrode patch setup can also be used instead of the band setup, however the positioning of each electrode must be chosen correctly.

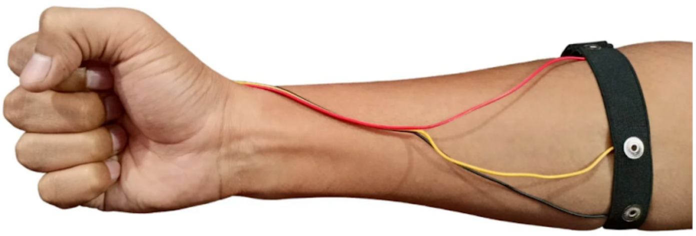


Figure 5.1: Single channel EMG acquisition band setup

5.2.2 Data Storage

Upon launching the MyoAuth application, users are presented with two options: Login or Register. For new users opting to register, they are prompted to enter a username. Following username registration, users undergo a 5-step EMG recording procedure. Each step involves a 4-second EMG recording session aimed at capturing specific muscle activity associated with a predefined gesture. The recorded EMG values, including user ID with attempt number, timestamps, and the corresponding EMG values, are then stored in a CSV file for further analysis and processing.

5.2.3 Signal Preprocessing

After data acquisition, the recorded EMG signals undergo preprocessing to eliminate unwanted signals or noise that may interfere with accurate analysis. Noises can be generated by :

- Electromagnetic interference: Interference from nearby electrical equipments and devices can induce unwanted electrical signals, which can corrupt the signals and generate noises

- Movement Artifacts: Movement of the electrode band or cables during the EMG recording can introduce motion artifacts, which generates noise. Muscle contraction or body movements can also cause noises.

- Muscle Cross Talk: When adjacent or overlapping muscles contaminate the signal,cross talk occurs. In multi channel EMG recordings, signals from neighboring muscles can leak into the recorded signal

- Electrode Impedance: Lack of skin preparation with electrode gel,size of the electrode and material can affect electrode impedance which results in poor electrical contact between electrode and skin. This results in noise.

Various measures can be taken to reduce noise. A user must clean his/her skin with alcohol for skin preparation and must use an electrode gel so as to avoid cross-talks from the muscle surface in contact. Furthermore, filters can be applied to filter out unwanted signals or noise from the raw signal. For this project, too mitigate noise, a high-pass Butterworth filter with a cutoff frequency of 5 Hz and notch filter with a cutoff frequency of 60 Hz is applied to the raw EMG data. This filter is selected because it effectively reduces noise while preserving essential signal components. The filtered EMG data is then saved to a separate CSV file, ready for subsequent segmentation and analysis.

5.2.4 Segmentation

During the registration or training phase, the preprocessed EMG signal, typically 4 seconds in duration, is segmented into three overlapping windows. Each window is 2 seconds long, with a 1-second overlap between consecutive windows. This segmentation technique results in 15 windows per user, derived from the 5 recorded signals segmented into 3 windows each.

In contrast, during the login or verification phase, a single 4-second signal is obtained from the user's registered gesture. This signal undergoes preprocessing similar to the registration phase and is then segmented into 41 overlapping windows. Each window is

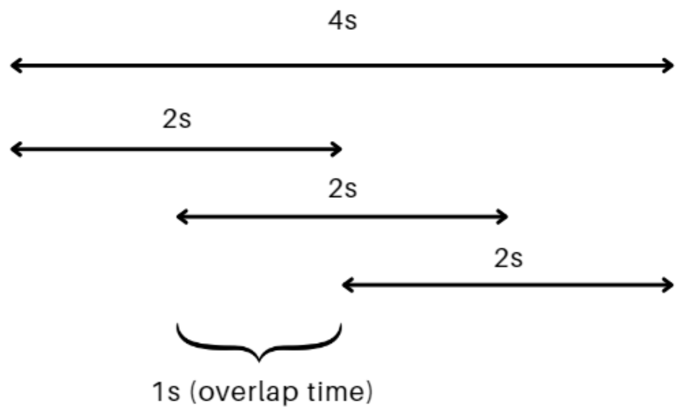


Figure 5.2: Segmentation during registration phase

2 seconds long, with a 50ms (0.05s) overlap between consecutive windows. This sliding window approach simplifies the comparison process during verification and ensures robustness against minor temporal variations in gesture execution.

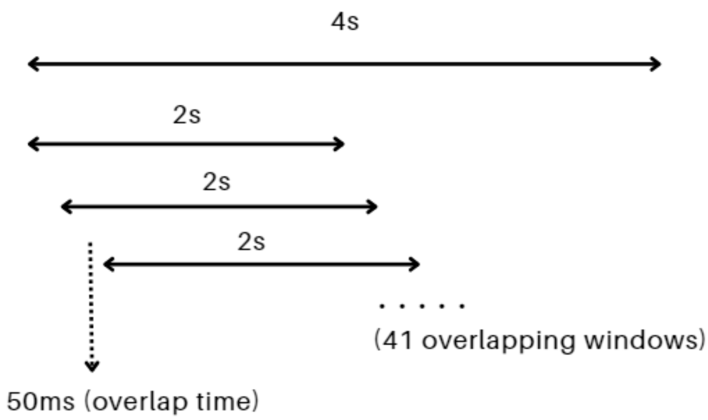


Figure 5.3: Segmentation during login phase

User Guidelines

Throughout the registration and verification phases, users are instructed to maintain consistent gesture execution in terms of power and rhythm. This consistency ensures accurate comparison and authentication, mitigating potential errors caused by minor variations in gesture performance. Additionally, users are encouraged to follow specific guidelines for electrode placement and skin preparation to optimize signal quality and minimize noise during EMG recording.

5.2.5 Feature Extraction

Following segmentation, features are extracted from each segmented window of the pre-processed signals. These features capture relevant information about muscle activity, such as amplitude, frequency, and temporal characteristics. For verification purposes, features extracted from the 41 windows of the login signal are compared with features from the 15 windows of the user's registered gesture. This comparison is facilitated using a trained K-nearest neighbors (KNN) model, enabling reliable authentication based on gesture similarity.

The following 21 time and frequency domain features were extracted from each segmented EMG signal and then combined to form a feature vector. Fast Fourier Transform was used to convert from time to frequency domain.

Mean Absolute Value (MAV):

MAV measures the average absolute magnitude of the EMG signal, indicating the overall muscle activity levels. It's calculated by taking the sum of the absolute values of all data points and dividing by the number of points:

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

Where x_i represents the EMG signal data points and N is the total number of data points.

Waveform Length (WL):

WL represents the cumulative length of the EMG waveform, capturing signal changes over time, useful for detecting muscle fatigue or contraction patterns. It's computed as the sum of the absolute differences between consecutive data points, normalized by a factor of 25:

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

Zero Crossing Rate (ZCR):

ZCR counts the number of times the EMG signal crosses the zero axis, indicating signal dynamics and rapid changes. It's calculated by counting the number of sign changes in the signal:

$$\text{ZCR} = \frac{\text{Number of Zero Crossings}}{N - 1}$$

Skewness:

Skewness measures the asymmetry of the EMG signal's distribution, which can indicate abnormal muscle activation patterns. The formula involves the mean $\bar{x}$ and calculates the third moment of the data divided by the cube of the standard deviation:

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{3/2}}$$

Kurtosis:

Kurtosis measures the peakedness or flatness of the EMG signal's distribution, helping detect sudden changes or anomalies. It's calculated by the fourth moment divided by the square of the standard deviation:

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{2}}$$

Root Mean Square (RMS):

RMS calculates the square root of the mean of the squared EMG signal values, providing a measure of signal magnitude and energy:

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

Simple Square Integral (SSI):

SSI represents the total energy content of the EMG signal, useful for assessing muscle activity intensity. It's computed as the sum of squared EMG signal values:

$$\text{SSI} = \sum_{i=1}^{N} x_i^2$$

Mean of Fourier Transform:

Computes the average frequency content of the EMG signal, helping identify dominant frequency components:

$$\text{Mean of FT} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Where X_i represents the Fourier transform coefficients of the EMG signal.

Frequency Centroid:

Calculates the center of mass of the frequency distribution, useful for characterizing spectral properties:

$$\text{Frequency Centroid} = \frac{\sum_{i=1}^{N} f_i X_i}{\sum_{i=1}^{N} X_i}$$

Where f_i is the frequency associated with X_i.

Total Energy:

Quantifies the overall energy of the EMG signal, important for assessing muscle activity levels and fatigue:

$$\text{Total Energy} = \sum_{i=1}^{N} X_i$$

Power:

Represents the signal power of the EMG waveform, useful for analyzing signal strength and intensity:

$$\text{Power} = \sum_{i=1}^{N} X_i^2$$

Dominant Frequency Index:

Identifies the index of the dominant frequency component in the EMG signal, useful for frequency-based analysis:

$$\text{Dominant Frequency Index} = \text{argmax}(X_i)$$

Mean Frequency:

Calculates the average frequency of the EMG signal, helping understand dominant frequency components:

$$\text{Mean Frequency} = \frac{\sum_{i=1}^{N} f_i X_i}{\sum_{i=1}^{N} X_i}$$

Median Frequency:

Determines the middle value of the frequency distribution, useful for identifying central frequency components:

$$\text{Median Frequency} = \text{Median}(f_i X_i)$$

Standard Deviation of Frequency:

Measures the variability of frequency components in the EMG signal. Useful for assessing signal stability.

$$\text{Std. Dev. of Frequency} = \text{Std. Dev.}(f_i X_i)$$

Variance of Frequency:

Quantifies the spread of frequency components in the EMG signal. Provides information about signal variability.

$$\text{Variance of Frequency} = \text{Variance}(f_i X_i)$$

Total Power (PSD):

Represents the total power contained in the EMG signal's power spectral density. Useful for overall power analysis.

$$\text{Total Power} = \sum_{i=1}^{N} P_i$$

Mean Power (PSD):

Calculates the average power in the power spectral density of the EMG signal. Useful for assessing signal strength.

$$\text{Mean Power} = \frac{1}{N} \sum_{i=1}^{N} P_i$$

Total Power Excluding DC Component (PSD):

Similar to total power but excludes the DC component. Useful for focusing on frequency-related power.

$$\text{Total Power Excl. DC} = \sum_{i=2}^{N} P_i$$

Mean Power Excluding DC Component (PSD):

Calculates the average power excluding the DC component in the power spectral density. Provides a more focused power measure.

$$\text{Mean Power Excl. DC} = \frac{1}{N-1} \sum_{i=2}^{N} P_i$$

Spectral Entropy:

Measures the randomness or complexity of the frequency components in the EMG signal. Useful for assessing signal diversity and information content.

$$\text{Spectral Entropy} = -\sum_{i=1}^{N} X_i \log_2(X_i + \epsilon)$$

5.2.6 Training

After extracting features, a set of fifteen feature vectors is generated for each user. To ensure optimal model performance, the dataset undergoes a meticulous partitioning process. Specifically, twelve feature vectors per user are designated for training purposes, while the remaining three vectors are earmarked for validation, thus creating a robust framework for model evaluation and refinement.

Utilizing the K-Nearest Neighbor (KNN) algorithm, configured with a parameter setting of three neighbors and employing the Manhattan distance as the primary distance metric, the model undergoes rigorous evaluation. KNN, renowned for its non-parametric nature applicable to both classification and regression tasks, functions by categorizing data points based on the predominant class of their closest neighbors. Through a meticulous computation of distances within the feature space, the algorithm efficiently assigns a class label to a query point by assessing the most prevalent class among its neighboring data points.

The Manhattan distance, D is given by:

$$\mathrm{D} = \sum_{i=1}^{n} |x_{2_i} - x_{1_i}|, \text{where } i \text{ ranges from 1 to } n.$$

During the critical phase of model evaluation, the system demonstrates an impressive validation accuracy rate of 0.83, signifying that a remarkable 83% of instances within the validation set are accurately classified. Additionally, the model's performance is further evaluated using the F1 score, a comprehensive metric combining precision and recall. The calculated F1 score of 0.8 highlights the model's effectiveness in capturing and managing both false positives and false negatives, showcasing its robustness and reliability in real-world authentication scenarios.

5.3 Description of Implementation Strategies

5.3.1 Libraries Used

- **SciPy:** contains modules used for signal processing like filters and fourier transforms
- **scikit-learn:** contains modules used for training
- **Tkinter and PIL:** contains modules used for UI design
- **matplotlib:** contains modules used for data visualiztion

5.3.2 Application of High Pass Filter and Notch Filter

The above code snippet defines two functions for filtering EMG (electromyography) data. The apply_highpass_filter function applies a high-pass filter with a cutoff frequency of 5 Hz to remove low-frequency noise from the input data, while the notch_filter function applies a notch (band-stop) filter at 60 Hz to eliminate interference from powerline noise. Both functions utilize functions from the scipy.signal module to design and apply digital filters. The high-pass filter's cutoff frequency is specified by the lowcut parameter, while the notch filter's center frequency is set to 60 Hz. These filters help preprocess EMG data by reducing noise and interference from unwanted frequencies, thus enhancing signal clarity for further analysis.

```python
def apply_highpass_filter(self, data, lowcut, fs=300.0, order=4):
    nyquist = 0.5 * fs
    low = lowcut / nyquist
    b, a = butter(order, low, btype='high')
    filtered_data = lfilter(b, a, data)
    return filtered_data

def notch_filter(self, data, f0, fs=300.0, Q=30.0):
    nyquist = 0.5 * fs
    normal_cutoff = f0 / nyquist
    b, a = iirfilter(2, [normal_cutoff - 1e-3, normal_cutoff + 1e-3], btype='bandstop')
    filtered_data = lfilter(b, a, data)
    return filtered_data
    #return filtered_amplitude
```

Figure 5.4: Code Snippet of Filtering

5.3.3 Segmentation of Filtered EMG Data

```
def segment_and_append_to_csv(self):
    # Define window length and overlap
    window_length = 2  # in seconds
    overlap = 1  # in seconds
    sample_rate = 300

    # Process each user's data separately
    unique_users = df['Username'].unique()
    for user_name in unique_users:
        user_df = df[df['Username'] == user_name]

        emg_data = user_df[['timestamps', 'emgvalues']].values.tolist()
        num_samples = len(emg_data)
        num_windows = int((num_samples - window_length * sample_rate) / (overlap * sample_rate)) + 1

        # Create a DataFrame to store segmented data
        columns = ["Username", "Window Number", "Start Time", "End Time", "emgvalues"]
        segmented_df = pd.DataFrame(columns=columns)

        for window_num in range(num_windows):
            start_index = int(window_num * overlap * sample_rate)
            end_index = int(start_index + window_length * sample_rate)

            window_data = emg_data[start_index:end_index]

            start_time = window_data[0][0]
            end_time = window_data[-1][0]
            emg_values = [emg_value for _, emg_value in window_data]

            # Append data to the DataFrame
            segmented_df.loc[len(segmented_df)] = [user_name, window_num + 1, start_time, end_time, emg_values]

        # Append segmented data to the CSV file
        if os.path.isfile(filename2):
            segmented_df.to_csv(filename2, mode='a', header=False, index=False)
        else:
            segmented_df.to_csv(filename2, index=False)
```

Figure 5.5: Code Snippet of Segmentation

The segment_and_append_to_csv function loads an existing CSV file containing filtered EMG (electromyography) data and segments it into fixed-size windows with a 2 second length and 1 second overlap. It iterates through each user's data, extracts the EMG values, and partitions them into consecutive windows. For each window, the function records the username, window number, start and end timestamps, and the corresponding EMG values. The segmented data is then appended to a new CSV file or an existing one if available, enabling easy access and analysis of segmented EMG data for further processing or modeling tasks.

5.3.4 Feature Extraction of the Segmented EMG Data

```
def extract_features(self, input_file, output_file):
    df = pd.read_csv(input_file)

    feature_vectors = []
    for index, row in df.iterrows():
        emg_values = eval(row['emgvalues'])  # Convert string to list
        features = self.calculate_features(emg_values)
        user_name = row['Username']
        feature_vectors.append({'Username': user_name[:-1], 'Features': features})

    feature_df = pd.DataFrame(feature_vectors)
    feature_df.to_csv(output_file, index=False)

    print(f"\nFeature vectors saved to {output_file}")
```

Figure 5.6: Code Snippet of Feature Extraction

The extract_features function reads EMG data from an input CSV file, iterates through each row to extract the EMG values, and calculates 21 features for each window using the calculate_features method. These features include time-domain and frequency-domain characteristics computed from the EMG signal. Subsequently, the features are combined into a feature vector along with the corresponding user name, and saved to an output CSV file. This process ensures that each window's EMG data is transformed into a concise representation suitable for further analysis or model training.

5.3.5 Model Training

The train_classifier method loads feature vectors from a CSV file, where each user possesses 15 feature vectors, and converts the string representations of features into lists. It then splits the data into training and testing datasets, with each user contributing three feature vectors to the testing dataset. The feature values are standardized to ensure uniform scaling across the dataset. Subsequently, a KNN classifier is instantiated and trained using the training data. The classifier utilizes a Manhattan distance metric and considers three nearest neighbors for classification. This process enables the model to learn the patterns present in the feature vectors and make predictions based on similarity to known instances, facilitating user authentication tasks effectively.

```python
def train_classifier(self):
    df = pd.read_csv(input_file)
    df['Features'] = df['Features'].apply(lambda x: eval(x))  # Convert string to list

    # 'Username' is the column containing user names
    unique_users = df['Username'].unique()

    # Create testing dataset with one feature vector per user
    testing_data = []
    for user in unique_users:
        user_data = df[df['Username'] == user].head(3)  # Select last 3 row for each user
        testing_data.append(user_data)

    testing_df = pd.concat(testing_data)
    training_df = df.drop(testing_df.index)

    # Extract features and labels from training and testing datasets
    X_train = np.vstack(training_df['Features'])
    y_train = training_df['Username']
    X_test = np.vstack(testing_df['Features'])
    y_test = testing_df['Username']

    print(y_test)

    # Standardize the feature values
    print("Standardizing feature values...")
    self.scaler = StandardScaler()
    X_train_scaled = self.scaler.fit_transform(X_train)
    X_test_scaled = self.scaler.transform(X_test)

    # Create and train the KNN classifier
    print("Training KNN classifier...")
    self.knn_classifier = KNeighborsClassifier(n_neighbors=3, metric='manhattan')
    self.knn_classifier.fit(X_train_scaled, y_train)
```

Figure 5.7: Code Snippet of Model Training

Chapter 6

Results and Discussions

6.1 Overview

In this section, the outcomes of the investigation into the development and evaluation of an authentication system leveraging electromyography (EMG) signals coupled with the K-Nearest Neighbors (KNN) algorithm are presented. The primary aim of this study was to assess the effectiveness and feasibility of utilizing EMG signals for user authentication purposes, with a specific focus on the performance of the KNN classifier in distinguishing between legitimate users and unauthorized individuals.

Throughout the experimentation phase, EMG data was collected from a group of participants performing a range of predefined hand gestures. Subsequently, various preprocessing techniques were applied to enhance the quality and reliability of the acquired signals, followed by feature extraction to capture relevant patterns indicative of individual muscle activity. Utilizing these features, the performance of the KNN classifier was trained and evaluated in differentiating between genuine users and impostors.

The analysis encompasses several key aspects, including the validation & training accuracy and F1 score of the authentication system across varying parameters such as the number of neighbors in the KNN algorithm and the dimensionality of the feature space. Furthermore, extensive experimentation was conducted to assess the robustness of the system against common challenges, such as noise, variability in signal acquisition conditions, and potential adversarial attacks.

6.2 Testing

Figure 6.1: Flash Page

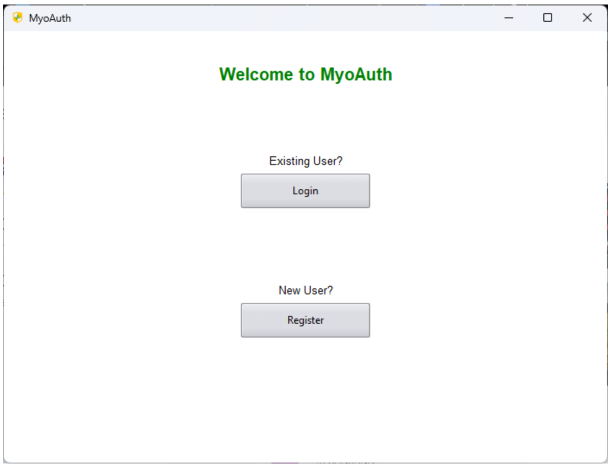


Figure 6.2: Landing Page

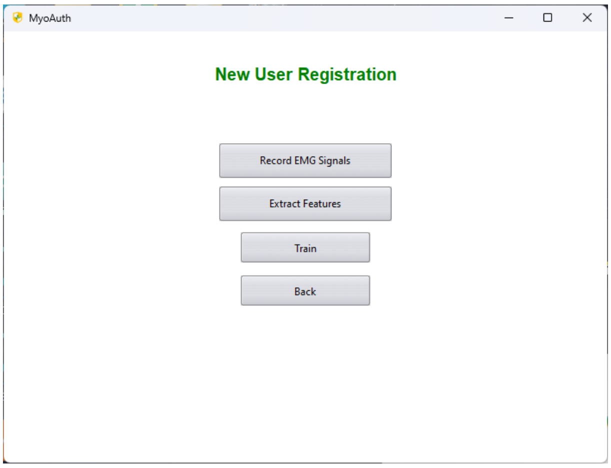


Figure 6.3: Registration Page

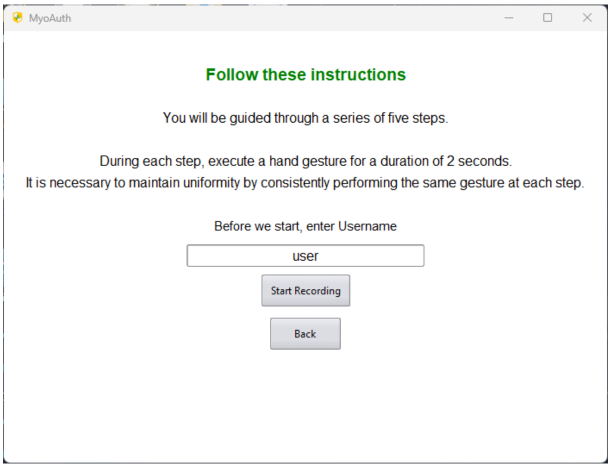


Figure 6.4: Registration Instructions Page

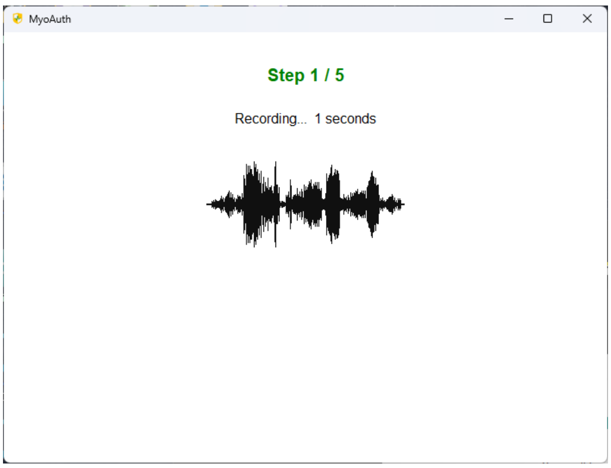


Figure 6.5: Signal Acquisition During Registration

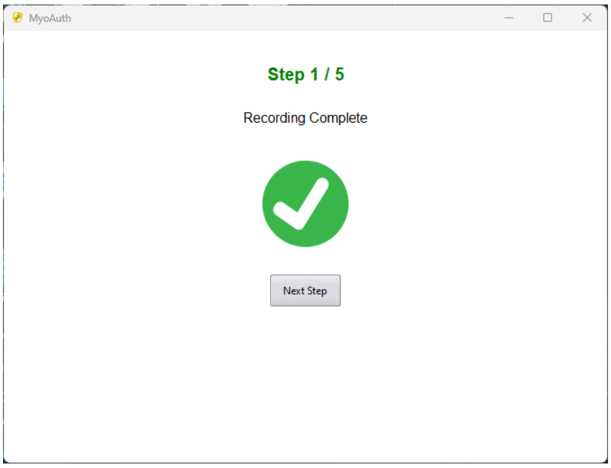


Figure 6.6: Signal Acquisition Complete

```
noel1,13:00:02.127,8.24
noel1,13:00:02.127,21.99
noel1,13:00:02.131,-14.09
noel1,13:00:02.131,-27.97
noel1,13:00:02.136,22.42
noel1,13:00:02.136,42.97
noel1,13:00:02.139,-11.96
noel1,13:00:02.139,-58.52
noel1,13:00:02.144,-26.95
noel1,13:00:02.144,51.45
noel1,13:00:02.147,76.49
noel1,13:00:02.148,-10.38
noel1,13:00:02.151,-107.66
noel1,13:00:02.152,-54.99
noel1,13:00:02.155,88.0
noel1,13:00:02.156,99.6
noel1,13:00:02.160,-22.2
noel1,13:00:02.160,-77.97
```

Figure 6.7: Raw Acquired Signal

```
Username,Window Number,Start Time,End Time,emgvalues
noel1,1,13:00:02.119,13:00:03.320,"[1.8058451916708385, 2.9477217316675, -6.520523453388197, -11.
91804441889464, 10.713843459529912, 20.023596087503183, -16.58574454138813, -24.278789050941164, 26.
03117001241835, 36.941183387513455, -20.91004741031011, -56.065206520387534, -13.759867798702563, 58.
15414621222234, 64.8987483144084, -27.252802594546637, -104.70286783080796, -32.33188489580276, 99.
56149282493097, 84.20396158515317, -41.399340764644286, -77.71990678825216, -16.557756418815877, 20.
60635463260615, 22.09642147460157, 23.32776635665316, 6.916491574697224, -21.295855930255215, -23.
524202394365737, -7.335685549398224, 5.423053167597137, 32.83386913912582, 42.05644490147339, -33.
299355341752175, -82.9665748385363, 14.804837440436051, 80.67626974673556, -23.52239895218529, -61.
98225770751189, 71.63039514905148, 75.34165256745864, -98.41082780284805, -96.03336601998316, 70.
07890370858016, 71.49580814993543, -29.306882904336, -16.183268617502126, 10.850760168006277, -26.
47543783230176, -5.92940786161556, 47.59584302641259, 8.37600851882802, -47.05991762412621, -12.
23578685396342, 28.932016656625127, -1.406994199128126, -27.23714805979056, 17.669353764237876, 49.
26684303134645, -5.657154517962805, -46.50575941590305, -7.061603373498388, 15.085598993657634, -4.
136525371589324, 0.7611976425752669, 3.8015593560321177, -3.0900147902620443, 18.552429770335568, 5.
368386911938972, -39.12862056799087, 3.1511287512798694, 61.13604789905285, -0.4546365873197358, -73.
36531550430288, -29.795747114198758, 40.32502875184072, 36.81684767559337, 5.13769328244043, -6.
034812858883815, -1.0757361841569928, 2.067487271971774, -27.402220727154678, -37.122532871682026, 29.
```

Figure 6.8: Filtered and Segmented Signal

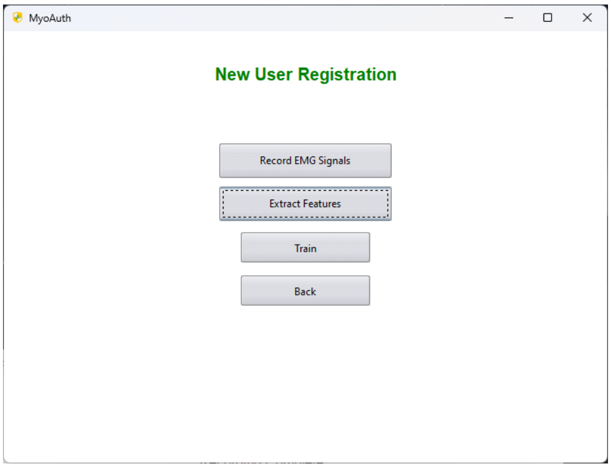


Figure 6.9: Feature Extraction on the Acquired Signal

```
Username,Features
noel,"[35.40843386082963, 26615.73356684042, 0.4373956594323873, -0.0443723323540481, 1.4189536149601398,
noel,"[36.06096880171826, 26799.497566019407, 0.4290484140233723, -0.040886683372161954, 1.23801046368192
noel,"[39.117827549206254, 28503.67707146225, 0.41903171953255425, 0.006938574945752152, 1.34318572673954
noel,"[37.53284403392347, 28860.99349444324, 0.4357262103505843, -0.04429336195949751, 1.226830453724122,
noel,"[38.826977959747566, 28726.34636965538, 0.41068447412353926, -0.05976567801001972, 0.79092499943323
noel,"[41.37796803085057, 30175.38586024975, 0.41903171953255425, -0.017693222190777572, 2.22498494387879
noel,"[39.479770800625396, 28575.89488052688, 0.41402337228714525, -0.06757490112144843, 3.19061044946433
noel,"[40.772476135315664, 29599.639902612053, 0.41569282136894825, -0.08854382645658189, 2.84267216153902
noel,"[37.69268050597075, 28101.460086944677, 0.41235392320534225, 0.008513762500806566, 1.85035504368475
noel,"[46.391379441548764, 35002.49564983901, 0.4357262103505843, 0.0025974918044310894, 1.58934375358787
noel,"[47.82210595095113, 35644.80539693526, 0.4290484140233723, -0.03088144090034503, 2.248150408267155,
noel,"[50.68820388937845, 38089.791514014054, 0.4240400667779633, -0.038873877257730546, 2.02865253283576
noel,"[37.460853222934674, 29059.960429296352, 0.44407345575959933, -0.017724879571871133, 2.0505292523343
noel,"[39.656005151979976, 30064.59786094525, 0.4340567612687813, -0.01127408120063211, 0.5779252554769019
noel,"[35.63584960555233, 26450.676138874027, 0.4373956594323873, -0.015274649994666226, 1.10715117586200
```

Figure 6.10: Extracted Features from the Segmented Signals

```
Training KNN classifier...
Training Accuracy: 0.875
Testing Accuracy: 0.8333333333333334
F1 Score: 0.8035714285714285
```

Figure 6.11: KNN model training results

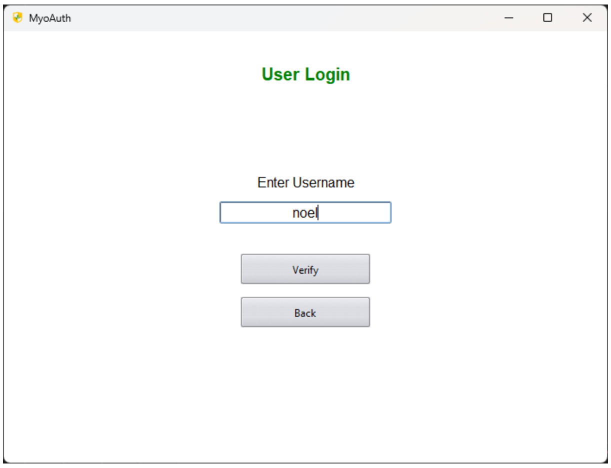


Figure 6.12: Login Page

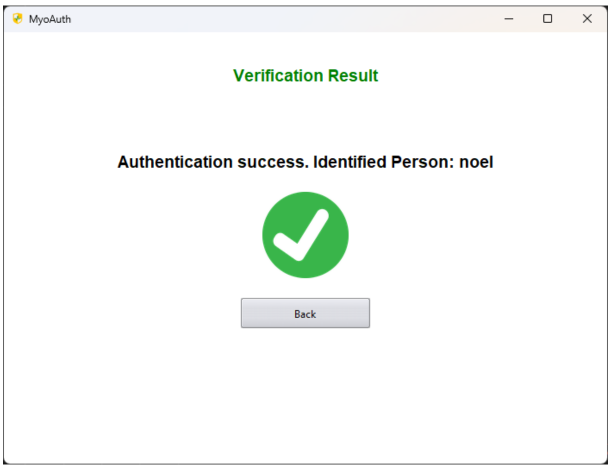


Figure 6.13: Login Result

6.3 Quantitative Results

The KNN model was trained using a dataset consisting of 4 users, with each user having a total of 15 feature vectors. From these fifteen feature vectors of a user, 12 were allocated for training and the remaining three were allocated for testing. During the training phase the model showed a training accuracy of 87.5%. During the testing phase, the model showed a testing accuracy of 83.3%.

In order to evaluate the model's predictive capability, F1 score was computed. This score which serves as a harmonic mean between precision and recall, was calculated to be 80%. F1 score provides a balanced evaluation of model's ability to correctly identify authentic(true positives) users while minimizing false positives and false negatives.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

During the login phase, where the user have to provide a new set of EMG signals and does the same set of movements provided during the registration phase, it was observed that most of the authentic users passed more than 15 windows out of 41 windows. This observation underscores the model's ability in accurately discerning genuine users from impostors, thereby showcasing its practical utility and relevance in real-world authentication scenarios.

6.4 Graphical Analysis

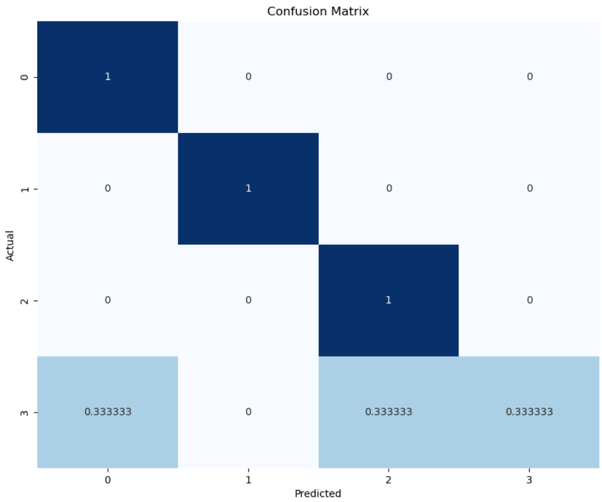


Figure 6.14: Confusion Matrix

The above confusion matrix is a tabular representation where each row corresponds to the actual classes or labels of the data, and each column represents the predicted classes assigned by the model. In this matrix, with four users in total, the confusion matrix dimension is 4x4.

Three feature vectors from each user were taken from the test dataset. The results were mapped between 0 and 1, representing the confidence scores model's predictions for each user.

The first three rows show perfect predictions for the corresponding users, indicating high accuracy in classification. However, the fourth row suggests a discrepancy, with predictions distributed more evenly across the classes, indicating lower accuracy for the fourth user. This suggests potential challenges or variability in the EMG signals associated with this user.

This matrix serves as a quantitative measure of the authentication system's efficacy, highlighting areas of success and potential improvement, particularly in ensuring consistent and accurate identification across all users.

6.5 Comparison with other models

While the K-Nearest Neighbors (KNN) model showed encouraging results with a limited dataset, exploring alternative models is essential to ensure a robust system.

Convolutional Neural Networks (CNNs) have proven effective for tasks like image and signal recognition. However, their success often relies heavily on having a very large training dataset. In this application, where user experience is also balanced, collecting a significant amount of EMG data using single channel during registration was not ideal.

Model	**Validation Accuracy**	**Suitability for Limited Dataset**	**Identification Accuracy**
K-Nearest Neighbors (KNN)	High (>70%)	Strong	High
Convolutional Neural Networks (CNN)	Low (<30%)	Weak	Low
One-Class Support Vector Machines (OCSVM)	Moderate (60%)	Moderate	Moderate

Table 2.6.4: Comparison of User Identification Models

An experiment was conducted where a CNN model was trained on the same dataset used for the KNN model. The CNN model achieved a validation accuracy below 30%, indicating its limitations with restricted data. Conversely, the KNN model's ability to function well with a smaller dataset makes it a more suitable option for this specific application.

One-Class Support Vector Machines (OCSVMs) were also considered as a potential approach, with each user having their own individual model. However, training these models for each user only yielded a maximum accuracy of 60%.

Based on these findings, the KNN model emerged as the preferred choice for achieving accurate user identification due to its strong performance with a limited dataset.

Chapter 7

Conclusion

MyoAuth, a pioneering authentication system, harnesses the power of electromyography (EMG) to introduce a robust alternative to current authentication technologies. By seamlessly integrating EMG data captured through Myo armbands and implementing an advanced k-NN classifier, MyoAuth offers heightened security through the authentication of unique muscle patterns. In the evolving landscape of authentication, MyoAuth stands out as a beacon of promise, providing a highly secure and reliable method for device authentication. What sets MyoAuth apart is its distinctive reliance on EMG data collected from Myo armbands, combined with the intelligence of the k-NN classifier. This synergy results in a sophisticated authentication system that adapts to the unique muscle signatures of users. MyoAuth ensures a smooth and intuitive authentication experience for users, eliminating the cumbersome processes associated with traditional authentication systems. Its robust performance is evidenced by its adaptability across diverse device configurations, accommodating variations in screen sizes, orientations, and placements. This innovative authentication system not only adds an extra layer of security but also positions itself as a valuable supplement to existing authentication mechanisms. MyoAuth demonstrates its potential to significantly enhance security across a broad spectrum of devices, making strides toward a future where user authentication is both reliable and user-friendly. As technology continues to advance, MyoAuth exemplifies a substantial leap forward in authentication technology, showcasing its applicability and adaptability in our increasingly interconnected digital world.

References

[1] Lu, L., Mao, J., Wang, W., Ding, G., Zhang, Z. (2020). A Study of Personal Recognition Method Based on EMG Signal. IEEE Transactions on Biomedical Circuits and Systems, 1–1. doi:10.1109/tbcas.2020.3005148

[2] Fan, B., Su, X., Niu, J., Hui, P. (2022). EmgAuth: Unlocking Smartphones with EMG Signals. IEEE Transactions on Mobile Computing,

[3] Raurale, S. A., McAllister, J., Rincon, J. M. D. (2021). "EMG Biometric Systems Based on Different Wrist-Hand Movements." IEEE Access, 9, 12256– 12266. doi:10.1109/access.2021.3050704

[4] P. Kang, S. Jiang and P. B. Shull, "Synthetic EMG Based on Adversarial Style Transfer Can Effectively Attack Biometric-Based Personal Identification Models," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 3275-3284, 2023, doi: 10.1109/TNSRE.2023.3303316.

[5] A. Pradhan, J. He and N. Jiang, "Performance Optimization of Surface Electromyography Based Biometric Sensing System for Both Verification and Identification," in IEEE Sensors Journal, vol. 21, no. 19, pp. 21718-21729, 1 Oct.1, 2021, doi: 10.1109/JSEN.2021.3079428.

[6] y. Xu, c. Yang, p. Liang, l. Zhao, and z. Li, development of a hybrid motion capturemethod using myo armband with applica- tion to teleoperation, in proc. Ieee int. Conf. Mechatron. Automat. , 2016, pp. 11791184.

[7] b. Fan, x. Liu, x. Su, p. Hui, and j. Niu, emgauth: an emg- based smartphoneunlocking system using siamese network, in proc. Ieee int. Conf. Pervasive comput.Commun. , 2020, pp. 110.

[8] a. Jain, l. Hong, and r. Bolle, on-line fingerprint verification, ieee trans. Patternanal.Mach. Intell. , vol. 19, no. 4, pp. 302314, apr. 1997.

[9] m. Sathiyanarayanan and s. Rajan, myo armband for physiother- apy healthcare:a case studyusing gesture recognition application, in proc. 8th int. Conf. Commun.Syst. Netw. , 2016, pp. 16.

[10] H. Yamaba et al., “An authentication method independent of tap operation on thetouchscreen of a mobile device,” J. Robot., Netw. Artif. Life, vol. 2, no. 1, pp. 60–63,2015.

[11] M. Pleva, Š. Korečko, D. Hladek, P. Bours, M. H. Skudal and Y. -F. Liao, ”Biometric User Identification by Forearm EMG Analysis,” 2022 IEEE International Conference on Consumer Electronics - Taiwan, Taipei, Taiwan, 2022, pp. 607-608, doi: 10.1109/ICCE-Taiwan55306.2022.9869268.

[12] De Luca, C. J. (1997). ”The use of surface electromyography in biomechanics.” J. Appl. Biomechanics, 13(2), 135-163.

[13] Greenberg, S. (2015). ”Myo Developer Blog.” Accessed: Feb. 22, 2018. https://developer.thalmic.com/forums/topic/1945/

[14] Sathiyanarayanan, M., Rajan, S. (2016). ”MYO armband for physiotherapy healthcare: A case study using gesture recognition application.” In Proc. 8th Int. Conf. Commun. Syst. Netw. (COMSNETS), Jan., pp. 16.

[15] Delsys. (2016). ”Delsys EMG Sensor Placement Technical Note 101.” Accessed: Jul. 9, 2018. https://www.delsys.com/downloads/TECHNICALNOTE/101-emgsensor-placement.pdf

[16] Raurale, S. A., McAllister, J., del Rincon, J. M. (2020). ”Real-time embedded EMG signal analysis for wrist-hand pose identification.” IEEE Trans. Signal Process., 68, 2713-2723.

[17] Ramírez Ángeles, I. J., Aceves Fernández, M. A. (2018). ”Multi-channel Electromyography Signal Acquisition of Forearm.” Available: https://data.mendeley.com/datasets/p77jn92bzg/1, doi: 10.17632/p77jn92bzg.1.

[18] Georgakis, A., Stergioulas, L. K., Giakas, G. (2003). "Fatigue analysis of the surface EMG signal in isometric constant force contractions using the averaged instantaneous frequency." IEEE Trans. Biomed. Eng., 50(2), 262-265.

[19] Phinyomark, A., Phukpattaranont, P., Limsakul, C. (2012). "Feature reduction and selection for EMG signal classification." Expert Syst. Appl., 39(8), 7420-7431.

[20] Klimesch, W., Schimke, H., Schwaiger, J. (1994). "Episodic and semantic memory: An analysis in the EEG theta and alpha band." Electroencephalogr. Clin. Neurophysiol., 91(6), 428-441.

Appendix A: Presentation

MyoAuth

Project Presentation

Team 8 - Guide: Ms. Dincy Paul

Nikhil Zachariah
Niya Bimal
Noel Joseph Paul
Sanjana Nair

May 10, 2024

Overview

1. Problem Definition
2. Project Objective
3. Novelty of Idea and Scope of Implementation
4. Literature Review
5. Methodology
6. Architecture Diagram
7. Sequence Diagram
8. Results
9. Work Division
10. Conclusion
11. Future Scope
12. References
13. Status of Paper Publication

Problem Definition

- Develop an EMG-based personal identification system that utilizes the distinct electromyographic patterns generated by individuals for accurate and secure identification.

Project Objective

- To create an enhanced biometric authentication system by implementing Electromyography(EMG) based Authentication.

Novelty of Idea and Scope of Implementation I

1. Unique Biometric Modality: Unlike traditional biometric methods such as fingerprints, iris scans, or facial recognition, EMG-based identification relies on capturing and analyzing the electrical activity produced by skeletal muscles during contraction. This makes it difficult to recreate an individual's EMG signal and thereby is a more reliable authentication method.

2. Voluntary Muscle Control/Liveness detection: EMG signals capture the electrical activity generated by voluntary muscle contractions initiated by the user. These contractions are under conscious control, allowing individuals to perform specific movements or gestures to authenticate their identity.

Novelty of Idea and Scope of Implementation II

3. Non-intrusive and User-Friendly: EMG-based authentication can be implemented using wearable devices such as armbands or wristbands embedded with sensors. These devices are non-intrusive and can be seamlessly integrated into daily routines, offering a user-friendly authentication experience without the need for explicit actions like scanning or tapping.

SCOPE:

1. Healthcare applications
2. Financial services authentication / Banking
3. Workplace access control
4. IoT device authentication & Smart home security
5. Government service integration

Literature Review

Paper	Description	Advantage	Disadvantage
[1]	Performance comparison of 2 methods: 1) Feature extraction methods : Hudgin's time domain, Frequency Division Technique (FDT), and Autoregressive Coefficients (AR) 2) Changing the no. of channels	Experimental conclusion: TD Feature Set : Offers superior authentication sensitivity and robustness. Four-channel Configuration : Balances performance and complexity effectively.	Limited Comparative Analysis: May overlook potential alternative methods. Empirical Validation: Requires further real-world testing for practical application.
[2]	Introduces an attack method based on generative strategies to generate synthetic EMG hand gesture signals, highlighting the potential risks of synthetic biological signals in identification systems.	Demonstrate the feasibility of the proposed method and highlight the vulnerability of deep EMG classifiers	The paper relies on the fact that EMG data is accessible to the attacker.

Literature Review

Paper	Description	Advantage	Disadvantage
[3]	The paper discusses related work on biometric authentication methods, EMG-based applications, and the Siamese network, and provides details on the EmgAuth system architecture	Learn the difference between a pair of inputs, which makes it well-suited for tasks that involve comparing EMG signals from different users.	Limited Comparative Analysis: May overlook potential alternative methods. Empirical Validation: Requires further real-world testing for practical application.
[4]	Acquire EMG signals from the user's wrist-hand movements, extracts time-domain features from the signals, and uses a Radial Basis Function Neural Network (RBF-NN) classifier to classify the features into different classes.	Simple to train and require fewer training samples compared to other types of neural networks	RBF-NNs can be computationally expensive, especially when dealing with large datasets or high-dimensional feature spaces

Methodology

1. Signal Acquisition

- EMG signals acquired using single-channel wearable band.
- Band includes three sensors, one on ulnar nerve for unique signal capture.
- Reliable signal acquisition ensures authentication accuracy.

2. Data Storage

- MyoAuth app offers Login or Register options.
- Registration involves 5-step EMG recording with user ID.
- Recorded data stored in CSV format for further analysis.

Methodology

3. Signal Preprocessing

- EMG signals undergo noise elimination.
- Noise sources include electromagnetic interference, movement artifacts, muscle cross talk, and electrode impedance.
- Measures like skin cleaning, gel usage, and filters applied for noise reduction.
- High pass filter (5 Hz) and notch filter (60 Hz) applied to eliminate noise.

Methodology

4. Segmentation (during registration phase)

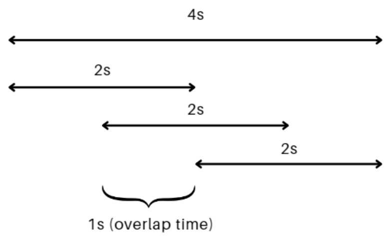


Figure: Segmentation during registration phase

- During the registration phase, the 4-second EMG signals are segmented into three overlapping windows. Each window is 2 seconds long, with a 1-second overlap between consecutive windows. This segmentation technique results in a total of 15 windows per user, derived from the 5 recorded signals segmented into 3 windows each.

Methodology

4. Segmentation (during login phase)

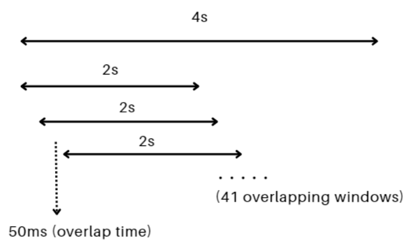


Figure: Segmentation during login phase

- During the login phase, a single 4-second signal is obtained from the user's registered gesture. This signal undergoes preprocessing similar to the registration phase and is then segmented into 41 overlapping windows. Each window is 2 seconds long, with a 50 ms (0.05 s) overlap between consecutive windows.

Feature Extraction

Fast Fourier Transform (FFT) is used to convert from time domain to frequency domain. Following 21 features of time and frequency domain are extracted from each segment.

1. Mean Absolute Value (MAV):

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

Measures average magnitude of EMG signal, indicating overall muscle activity levels.

2. Waveform Length (WL):

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

Represents the cumulative length of the EMG waveform, capturing signal changes over time.

Feature Extraction (contd.)

3. Zero Crossing Rate (ZCR):

$$\text{ZCR} = \frac{\text{Number of Zero Crossings}}{N-1}$$

Counts the rate at which the EMG signal crosses the zero axis, indicating signal dynamics.

4. Skewness:

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{3/2}}$$

Measures the asymmetry of the EMG signal's distribution, useful for detecting abnormal muscle activation patterns.

Feature Extraction (contd.)

5. Kurtosis:

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^2}$$

Measures the peakedness or flatness of the EMG signal's distribution, helping detect sudden changes or anomalies.

6. Root Mean Square (RMS):

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

Calculates the square root of the mean of the squared EMG signal values, providing a measure of signal magnitude and energy.

Feature Extraction (contd.)

7. Simple Square Integral (SSI):

$$\text{SSI} = \sum_{i=1}^{N} x_i^2$$

Represents the total energy content of the EMG signal, useful for assessing muscle activity intensity.

8. Mean of Fourier Transform:

$$\text{Mean of FT} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Computes the average frequency content of the EMG signal, helping identify dominant frequency components.

Feature Extraction (contd.)

9. Frequency Centroid:

$$\text{Frequency Centroid} = \frac{\sum_{i=1}^{N} f_i X_i}{\sum_{i=1}^{N} X_i}$$

Calculates the center of mass of the frequency distribution, useful for characterizing spectral properties.

10. Total Energy:

$$\text{Total Energy} = \sum_{i=1}^{N} X_i$$

Quantifies the overall energy of the EMG signal, important for assessing muscle activity levels and fatigue.

Feature Extraction (contd.)

11. Power:

$$\text{Power} = \sum_{i=1}^{N} X_i^2$$

Represents the signal power of the EMG waveform, useful for analyzing signal strength and intensity.

12. Dominant Frequency Index:

$$\text{Dominant Frequency Index} = \text{argmax}(X_i)$$

Identifies the index of the dominant frequency component in the EMG signal, useful for frequency-based analysis.

Feature Extraction (contd.)

13. Mean Frequency:

$$\text{Mean Frequency} = \frac{\sum_{i=1}^{N} f_i X_i}{\sum_{i=1}^{N} X_i}$$

Calculates the average frequency of the EMG signal, helping understand dominant frequency components.

14. Median Frequency:

$$\text{Median Frequency} = \text{Median}(f_i X_i)$$

Determines the middle value of the frequency distribution, useful for identifying central frequency components.

Feature Extraction (contd.)

15. Standard Deviation of Frequency:

$$\text{Std. Dev. of Frequency} = \text{Std. Dev.}(f_i X_i)$$

Measures the variability of frequency components in the EMG signal, useful for assessing signal stability.

16. Variance of Frequency:

$$\text{Variance of Frequency} = \text{Variance}(f_i X_i)$$

Quantifies the spread of frequency components in the EMG signal, providing information about signal variability.

Feature Extraction (contd.)

17. Total Power (PSD):

$$\text{Total Power} = \sum_{i=1}^{N} P_i$$

Represents the total power contained in the EMG signal's power spectral density.

18. Mean Power (PSD):

$$\text{Mean Power} = \frac{1}{N} \sum_{i=1}^{N} P_i$$

Calculates the average power in the power spectral density of the EMG signal.

Feature Extraction (contd.)

19. Total Power Excluding DC Component (PSD):

$$\text{Total Power Excl. DC} = \sum_{i=2}^{N} P_i$$

Similar to total power but excludes the DC component, focusing on frequency-related power.

20. Mean Power Excluding DC Component (PSD):

$$\text{Mean Power Excl. DC} = \frac{1}{N-1} \sum_{i=2}^{N} P_i$$

Calculates the average power excluding the DC component in the power spectral density.

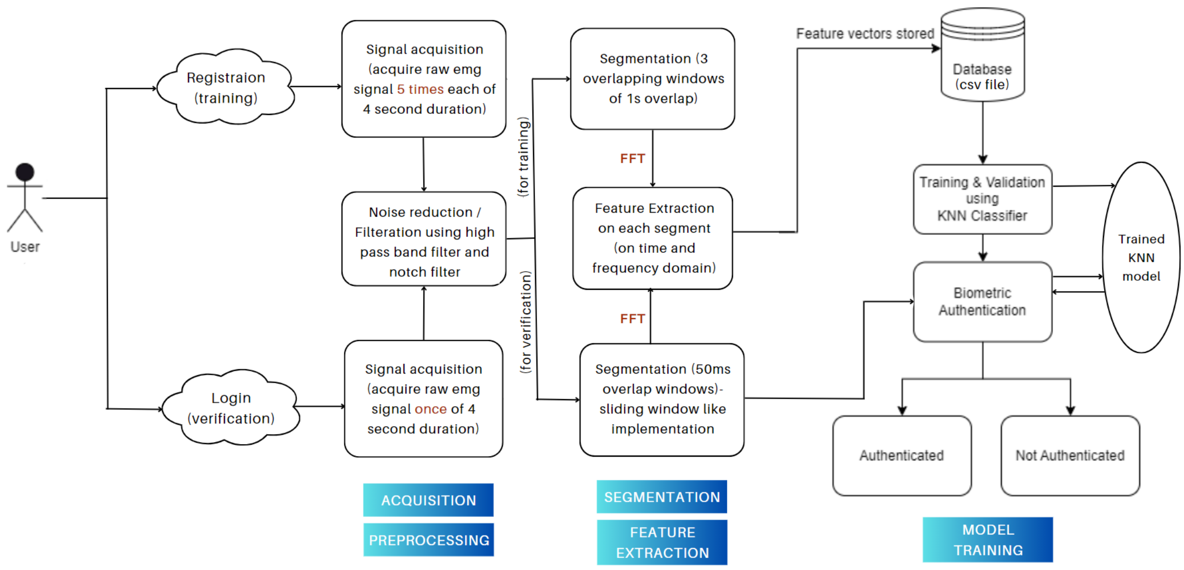
Feature Extraction (contd.)

21. Spectral Entropy:

$$\text{Spectral Entropy} = -\sum_{i=1}^{N} X_i \log_2(X_i + \epsilon)$$

Measures the randomness or complexity of the frequency components in the EMG signal.

Architecture Diagram



Sequence Diagram

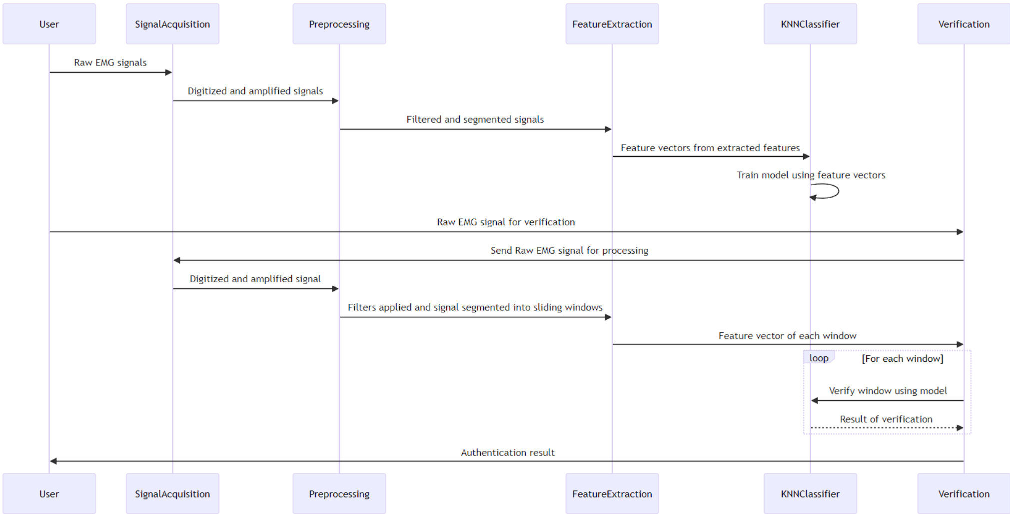


Figure: Sequence Diagram

Sequence Diagram

Registration Phase:

- Raw EMG signals acquired from user wearing band.
- Signals digitized, amplified, and sent to preprocessing module.
- High pass filter (5 Hz) and notch filter (60 Hz) applied to eliminate noise.
- Data segmented to isolate relevant muscle activity.
- Twenty-one features extracted in time and frequency domain.
- Feature vectors created for machine learning model input.
- Labeled feature vectors used to train KNN model.

Sequence Diagram

Login Phase:

- User records raw EMG signal.
- Signal digitized for efficient processing.
- Amplification enhances signal detection.
- Filters applied and signal segmented into sliding windows.
- Feature vectors extracted from each window.
- Verification done using trained KNN model.
- Authentication result sent to user.

Training of Model

- Fifteen feature vectors generated per user post feature extraction.
- Dataset partitioned:
  - Twelve feature vectors per user designated for training.
  - Three feature vectors per user were used for validation.
- K-Nearest Neighbor (KNN) algorithm utilized:
  - Configured with three neighbors.
  - Manhattan distance employed as primary distance metric.
  - Achieved impressive validation accuracy rate of 0.83 (83% accuracy).
  - F1 score calculated to be 0.8, highlighting precision and recall balance.
- Model showcases robustness and reliability in real-world authentication scenarios.

Results

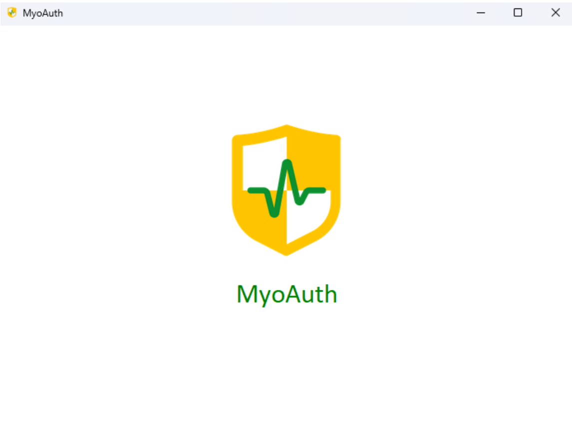


Figure: Splash Screen

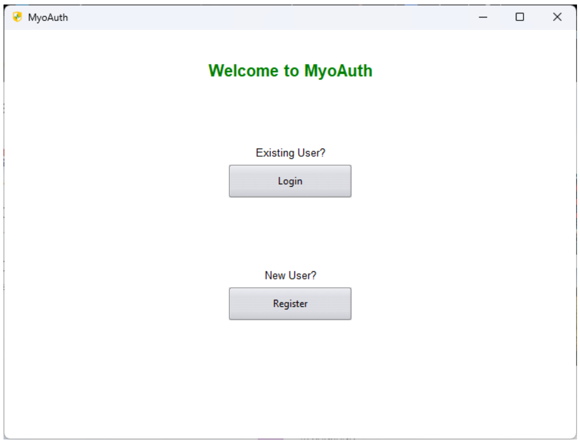


Figure: Home Page

Results

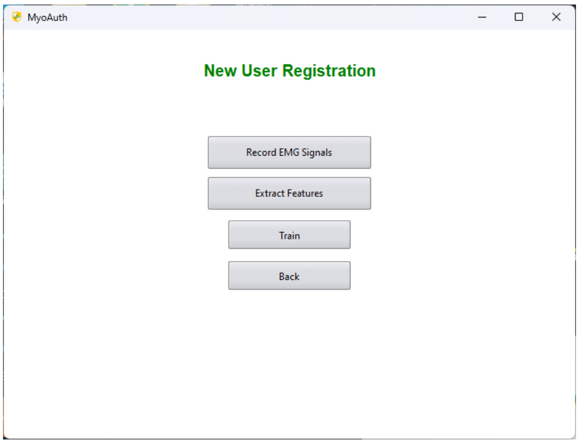


Figure: User Registration Home Page for User

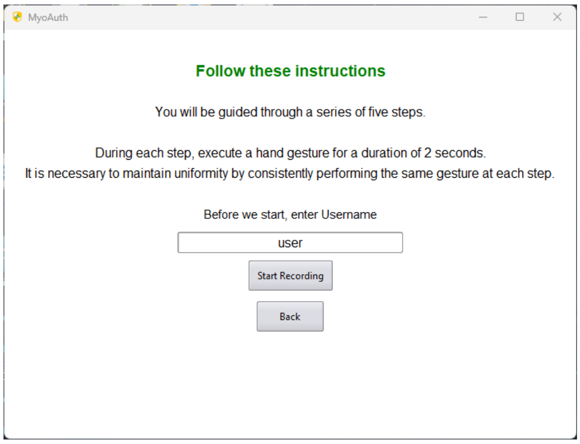


Figure: Instructions

Results

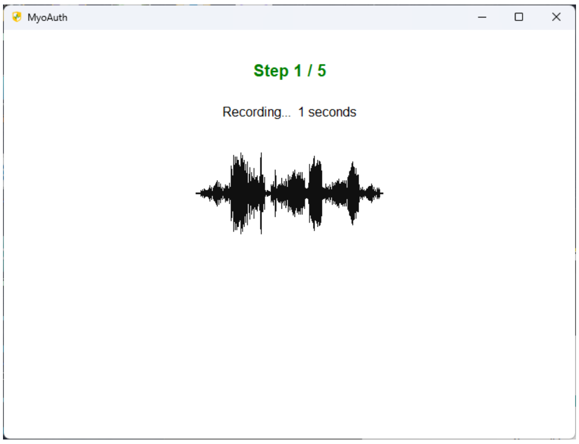
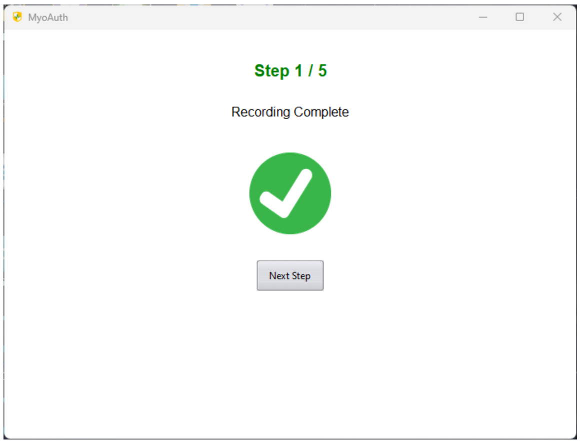


Figure: EMG Recording



Step Complete Message

Results

```
noel1,13:00:02.127,8.24
noel1,13:00:02.127,21.99
noel1,13:00:02.131,-14.09
noel1,13:00:02.131,-27.97
noel1,13:00:02.136,22.42
noel1,13:00:02.136,42.97
noel1,13:00:02.139,-11.96
noel1,13:00:02.139,-58.52
noel1,13:00:02.144,-26.95
noel1,13:00:02.144,51.45
noel1,13:00:02.147,76.49
noel1,13:00:02.148,-10.38
noel1,13:00:02.151,-107.66
noel1,13:00:02.152,-54.99
noel1,13:00:02.155,88.0
noel1,13:00:02.156,99.6
noel1,13:00:02.160,-22.2
noel1,13:00:02.160,-77.97
```

Figure: Raw EMG Data

```
Username,timestamps,emgvalues
noel1,13:00:02.119,1.8058451916708385
noel1,13:00:02.119,2.9477217316675
noel1,13:00:02.124,-6.520523453388197
noel1,13:00:02.124,-11.91804441889464
noel1,13:00:02.127,10.713843459529912
noel1,13:00:02.127,20.023596087503186
noel1,13:00:02.131,-16.58574454138813
noel1,13:00:02.131,-24.278789050941164
noel1,13:00:02.136,26.03117001241835
noel1,13:00:02.136,36.941183387513455
noel1,13:00:02.139,-20.91004741031011
noel1,13:00:02.139,-56.065206520387534
noel1,13:00:02.144,-13.759867798702565
noel1,13:00:02.144,58.154146212222344
noel1,13:00:02.147,64.8987483144084
noel1,13:00:02.148,-27.252802594546633
noel1,13:00:02.151,-104.70286783080795
```

EMG Data after applying Filters

Results

```
Username,Window Number,Start Time,End Time,emgvalues
noel1,1,13:00:02.119,13:00:03.320,"[1.8058451916708385, 2.9477217316675, -6.520523453388197, -11.
91804441889464, 10.713843459529912, 20.023596087503183, -16.58574454138813, -24.278789050941164, 26.
03117001241835, 36.941183387513455, -20.91004741031011, -56.065206520387534, -13.759867798702563, 58.
15414621222234, 64.8987483144084, -27.25280259454663\7, -104.70286783080796, -32.33188489580276, 99.
56149282493097, 84.20396158515317, -41.399340764644286, -77.71990678825216, -16.557756418815877, 20.
60635463260615, 22.09642147460157, 23.32776635665316, 6.916491574697224, -21.295855930255215, -23.
524202394365737, -7.335685549398224, 5.423053167597137, 32.83386913912582, 42.05644490147339, -33.
299355341752175, -82.9665748385363, 14.804837440436051, 80.67626974673556, -23.52239895218529, -61.
98225770751189, 71.63039514905148, 75.34165256745864, -98.41082780284805, -96.03336601998316, 70.
07890370858016, 71.49580814993543, -29.306882904336, -16.183268617502126, 10.850760168006277, -26.
47543783230176, -5.92940786161556, 47.59584302641259, 8.37600851882802, -47.05991762412621, -12.
23578685396342, 28.932016656625127, -1.406994199128126, -27.23714805979056, 17.669353764237876, 49.
26684303134645, -5.657154517962805, -46.50575941590305, -7.061603373498388, 15.085598993657634, -4.
136525371589324, 0.7611976425752669, 3.8015593560321177, -3.0900147902620443, 18.552429770335568, 5.
368386911938972, -39.12862056799087, 3.1511287512798694, 61.13604789905285, -0.4546365873197358, -73.
36531550430288, -29.795747114198758, 40.32502875184072, 36.81684767559337, 5.13769328244043, -6.
034812858883815, -1.0757361841569928, 2.067487271971774, -27.402220727154678, -37.122532871682026, 29.
```

Figure: Filtered and Segmented Signals

Results

```
segmented_emg_data.csv > data
Username,Window Number,Start Time,End Time,emgvalues
niya1,1,12:57:29.639,12:57:30.839,"[0.95, 1.37, -3.0, -4.61, 6.44, 11.33, -12.8, -28.12, 23.61, 65.82, -28.18, -123.29, -6.37, 148.09, 68.74, -91
niya1,2,12:57:30.242,12:57:31.442,"[-64.62, -31.22, 19.81, 26.56, -1.77, -4.05, 17.63, 0.98, -28.96, -5.62, 18.38, -14.68, -4.41, 63.98, -0.19, -
niya1,3,12:57:30.839,12:57:32.040,"[5.61, -148.06, -83.61, 96.57, 105.86, -0.19, -41.43, -36.53, -29.1, 8.25, 40.86, 21.65, -9.59, -14.87, -12.53
niya2,1,12:57:38.012,12:57:39.212,"[14.72, 29.69, -42.79, -125.54, 17.23, 211.05, 67.27, -188.42, -101.95, 120.25, 65.75, -88.05, -32.54, 67.56,
niya2,2,12:57:38.611,12:57:39.811,"[4.46, -6.44, -12.05, 1.96, 9.22, 0.58, -2.69, 1.69, -0.05, -3.99, -10.29, -15.13, 28.71, 70.53, -20.54, -111.
niya2,3,12:57:39.212,12:57:40.412,"[-25.29, -21.76, 3.7, 15.58, 7.81, 2.59, -0.97, -8.81, -7.77, 3.0, 6.99, 2.56, -0.44, -14.1, -29.69, 31.79, 11
niya3,1,12:57:46.516,12:57:47.716,"[13.79, 22.99, -46.49, -94.37, 52.55, 156.4, -14.52, -131.98, -8.58, 53.46, -12.75, 0.1, 40.73, -7.93, -39.83,
niya3,2,12:57:47.115,12:57:48.315,"[19.97, 20.31, 10.75, -26.36, -42.21, 15.69, 61.78, 5.07, -75.16, -38.11, 78.48, 60.25, -75.19, -54.95, 73.53,
niya3,3,12:57:47.716,12:57:48.917,"[91.89, -4.2, -24.39, -28.44, -42.51, 0.98, 45.51, 28.98, 0.6, -17.33, -33.68, -17.66, 22.64, 38.57, 18.13, -2
niya4,1,12:57:53.845,12:57:55.044,"[14.74, 29.76, -42.91, -125.96, 17.57, 212.18, 75.97, -178.44, -144.57, 56.7, 141.26, 63.4, -81.34, -122.04, -
niya4,2,12:57:54.443,12:57:55.642,"[-5.24, -5.44, 1.71, 3.66, 1.41, 0.97, -1.01, -9.73, -17.38, 10.24, 60.84, 31.96, -73.38, -86.45, 18.42, 69.97
niya4,3,12:57:55.045,12:57:56.245,"[5.06, -2.05, -8.74, -4.77, 2.02, 3.3, 3.27, 2.54, -5.36, -20.52, -9.54, 53.87, 69.61, -50.91, -129.86, -10.41
niya5,1,12:58:01.103,12:58:02.303,"[3.64, 6.82, -13.06, -33.27, 10.46, 66.01, 21.28, -64.22, -56.59, 21.41, 57.07, 18.87, -26.58, -22.81, -0.63,
niya5,2,12:58:01.706,12:58:02.905,"[0.23, 2.12, -1.57, -21.2, -10.03, 63.2, 71.21, -69.38, -138.62, 5.53, 117.99, 42.96, -31.05, -10.39, -17.0, -
niya5,3,12:58:02.307,12:58:03.503,"[2.85, 1.15, -1.32, -3.25, -9.07, -15.16, 19.22, 62.33, 1.96, -87.52, -38.12, 52.26, 32.64, -5.16, 9.32, -3.56
nikhil1,1,12:58:37.805,12:58:39.006,"[8.63, 23.24, -15.91, -96.61, -30.25, 162.24, 113.89, -158.79, -147.94, 135.27, 139.48, -114.22, -128.94, 61
nikhil1,2,12:58:38.407,12:58:39.608,"[122.97, 75.76, -80.14, -96.15, -0.27, 40.76, 39.32, 70.82, 65.07, -72.52, -186.07, -59.9, 148.0, 143.97, -7
nikhil1,3,12:58:39.009,12:58:40.210,"[65.6, -18.15, -100.27, -47.83, 104.14, 123.51, -41.67, -123.74, 3.09, 91.77, -2.1, -76.08, -32.65, 7.2, 50.
nikhil2,1,12:58:45.326,12:58:46.526,"[6.31, 10.56, -21.78, -42.44, 35.96, 83.59, -43.11, -120.96, 30.48, 134.31, 7.51, -94.44, -34.16, 21.69, 26.
nikhil2,2,12:58:45.928,12:58:47.129,"[33.5, -94.0, -38.89, 46.79, 0.64, -65.66, -11.64, 94.62, 77.38, -52.97, -107.53, -28.94, 72.03, 104.44, 23.
nikhil2,3,12:58:46.530,12:58:47.726,"[-50.12, 4.13, 16.71, -27.8, -20.48, -3.15, 8.51, 66.77, 56.4, -76.6, -98.89, 63.34, 117.79, -48.81, -160.38
nikhil3,1,12:58:54.269,12:58:55.468,"[14.63, 28.11, -50.17, -134.22, 36.66, 266.94, 89.01, -267.39, -225.42, 92.48, 198.91, 77.32, -22.68, -62.49
```

Figure: Segmented EMG Data file

Results

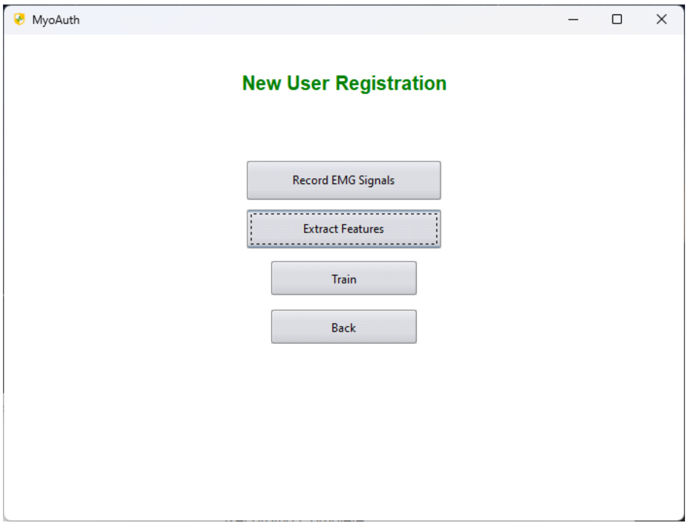


Figure: Feature Extraction

Results

```
Username,Features
noel,"[35.40843386082963, 26615.73356684042, 0.4373956594323873, -0.0443723323540481, 1.4189536149601398,
noel,"[36.06096880171826, 26799.497566019407, 0.4290484140233723, -0.040886683372161954, 1.23801046368192
noel,"[39.117827549206254, 28503.67707146225, 0.41903171953255425, 0.006938574945752152, 1.343185726739544
noel,"[37.53284403392347, 28860.99349444324, 0.4357262103505843, -0.04429336195949751, 1.22683045372412,
noel,"[38.826977959747566, 28726.34636965538, 0.41068447412353926, -0.05976567801001972, 0.790924999433234
noel,"[41.37796803085057, 30175.38586024975, 0.41903171953255425, -0.017693222190777572, 2.22498494387879
noel,"[39.479770800625396, 28575.89488052688, 0.41402337228714525, -0.06757490112144843, 3.190610449464334
noel,"[40.772476135315664, 29599.639902612053, 0.41569282136894825, -0.08854382645658189, 2.84267216153902
noel,"[37.69268050597075, 28101.460086944677, 0.41235392320534225, 0.008513762500806566, 1.850355043684750
noel,"[46.391379441548764, 35002.49564983901, 0.4357262103505843, 0.0025974918044310894, 1.589343753587879
noel,"[47.82210595095113, 35644.80539693526, 0.4290484140233723, -0.03088144090034503, 2.248150408267155,
noel,"[50.68820388937845, 38089.791514014054, 0.4240400667779633, -0.03887387725773054б, 2.028652532835760
noel,"[37.460853222934674, 29059.960429296352, 0.44407345575959933, -0.017724879571871133, 2.050529252334З
noel,"[39.656005151979976, 30064.59786094525, 0.4340567612687813, -0.01127408120063211, 0.577925255476901
noel,"[35.63584960555233, 26450.676138874027, 0.4373956594323873, -0.015274649994666226, 1.10715117586200
```

Figure: Extracted Features

```
Training KNN classifier...
Training Accuracy: 0.875
Testing Accuracy: 0.8333333333333334
F1 Score: 0.8035714285714285
```

Figure: KNN model training results

Results

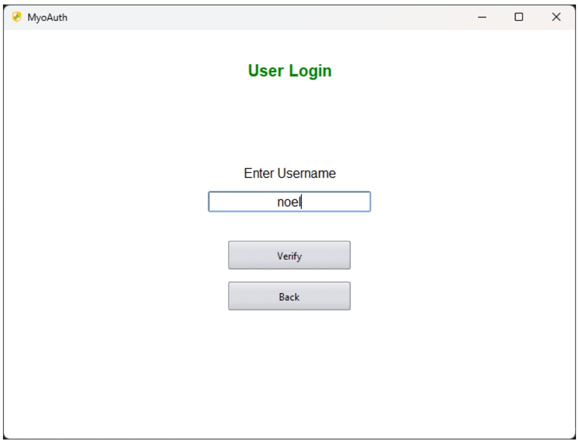
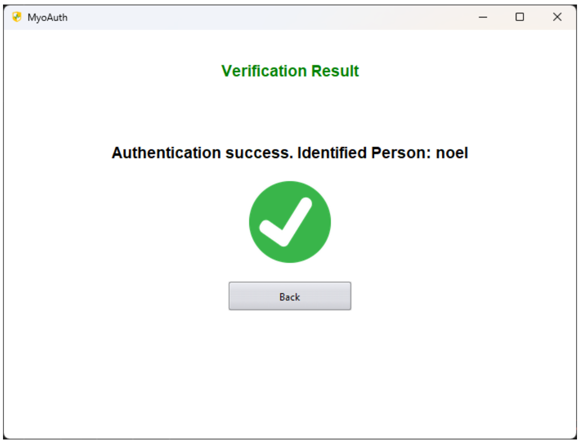


Figure: User Login Page



User Authenticated

Future Scope

- **Multi-Channel EMG Systems:** Increasing the number of channels in EMG systems can capture muscle activity from multiple locations, providing a more comprehensive representation of muscle activation patterns.

- **Expansion of Training Dataset:** Incorporating EMG data from a broader range of individuals, demographics, and activities can improve the classifier's ability to recognize unique muscle activation patterns, leading to better authentication performance across various users and scenarios.

Task Distribution (100%)

1. **Nikhil Zachariah:**
   -Focused on understanding and handling EMG sensors/devices.
   -Managed signal acquisition and transmission processes.
   -Applied filters on raw EMG for noise reduction purposes
   -Contributed to UI development

2. **Niya Bimal:**
   -Performed segmentation of raw EMG data into overlapping windows in training and validation phases.
   -Performed feature extraction of various frequency domain features on each of the segmented windows.
   -Contributed to UI development

Task Distribution (100%)

3. **Noel Joseph Paul:**
   - Implemented KNN classifier to train the acquired feature sets
   - Fine-tuned parameters to improve model accuracy
   - Contributed to UI development

4. **Sanjana Nair:**
   - Applied Fast Fourier Transform to convert from time domain to frequency domain
   - Performed feature extraction of various time domain features on each of the segmented windows.
   - Contributed to UI development

Conclusion

- Developed an authentication system utilizing electromyography (EMG) signals, and leveraging features.
- Trained a machine learning classifier, based on a dataset of labeled EMG feature vectors, to distinguish between users' unique muscle activation patterns.
- Implemented feature verification, comparing recorded EMG signals to trained patterns to authenticate users, providing a potentially more secure authentication method than traditional password-based systems.

References

[1] Fan, B., Su, X., Niu, J., Hui, P. (2022). EmgAuth: Unlocking Smartphones with EMG Signals. IEEE Transactions on Mobile Computing,
[2] P. Kang, S. Jiang and P. B. Shull, "Synthetic EMG Based on Adversarial Style Transfer Can Effectively Attack Biometric-Based Personal Identification Models," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 3275-3284, 2023, doi: 10.1109/TNSRE.2023.3303316.
[3] Fan, B., Su, X., Niu, J., Hui, P. (2022). EmgAuth: Unlocking Smartphones with EMG Signals. IEEE Transactions on Mobile Computing,
[4] Raurale, S. A., McAllister, J., Rincon, J. M. D. (2021). "EMG Biometric Systems Based on Different Wrist-Hand Movements." IEEE Access, 9, 12256- 12266. doi:10.1109/access.2021.3050704

References

[5] M. Pleva, Š. Korečko, D. Hladek, P. Bours, M. H. Skudal and Y. -F. Liao, ”Biometric User Identification by Forearm EMG Analysis,” 2022 IEEE International Conference on Consumer Electronics - Taiwan, Taipei, Taiwan, 2022, pp. 607-608, doi: 10.1109/ICCE-Taiwan55306.2022.9869268.
[6] Lu, L., Mao, J., Wang, W., Ding, G., Zhang, Z. (2020). A Study of Personal Recognition Method Based on EMG Signal. IEEE Transactions on Biomedical Circuits and Systems, 1–1. doi:10.1109/tbcas.2020.3005148
[7] A. Pradhan, J. He and N. Jiang, ”Performance Optimization of Surface Electromyography Based Biometric Sensing System for Both Verification and Identification,” in IEEE Sensors Journal, vol. 21, no. 19, pp. 21718-21729, 1 Oct.1, 2021, doi: 10.1109/JSEN.2021.3079428.

References

[8] a. Jain, l. Hong, and r. Bolle, on-line fingerprint verification, ieee trans. Patternanal.Mach. Intell. , vol. 19, no. 4, pp. 302314, apr. 1997.
[9] m. Sathiyanarayanan and s. Rajan, myo armband for physiother- apy healthcare:a case studyusing gesture recognition application, in proc. 8th int. Conf. Commun.Syst. Netw. , 2016, pp. 16.
[10] H. Yamaba et al., “An authentication method independent of tap operation on thetouchscreen of a mobile device,” J. Robot., Netw. Artif. Life, vol. 2, no. 1, pp. 60–63,2015.

Status of Paper Publication

Conference: 2024 International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)

Submission Date: May 10, 2024

Conference Date: 20th – 22nd September 2024

Venue: Indian Institute of Information Technology Kottayam, Kerala, India

Paper publisher: IEEE

Achievements

Conference: 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Venue: COEX, Seoul, Korea

Result: Accepted for Show and Tell Proposal

Thank you

Team 8 – Guide: Ms. Dincy Paul

Nikhil Zachariah
Niya Bimal
Noel Joseph Paul
Sanjana Nair

May 10, 2024

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision
To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission
To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision
To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

Department Mission
To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

Programme Outcomes (PO)
Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team work: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

10. Communication: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)
A graduate of the Computer Science and Engineering Program will demonstrate:

PSO1: Computer Science Specific Skills
The ability to identify, analyze and design solutions for complex engineering problems

in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

PSO2: Programming and Software Development Skills
The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

PSO3: Professional Skills
The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

Course Outcomes(CO)

SNO	DESCRIPTION
CO1	Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).
CO2	Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).
CO3	Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).
CO4	Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).
CO5	Identify technology/research gaps and propose innovative/cre- ative solutions (Cognitive knowledge level: Analyze).
CO6	Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

Appendix C: CO-PO-PSO Mapping

CO-PO AND CO-PSO MAPPING

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO 1	2	2	2	1	2	2	2	1	1	1	1	2	3		
CO 2	2	2	2		1	3	3	1	1		1	1		2	
CO 3									3	2	2	1			3
CO 4					2			3	2	2	3	2			3
CO 5	2	3	3	1	2							1	3		
CO 6					2			2	2	3	1	1			3

3/2/1: high/medium/low

JUSTIFICATIONS FOR CO-PO MAPPING

MAPPING	LOW/MEDIUM/ HIGH	JUSTIFICATION
100003/ CS722U.1-PO1	M	Knowledge in the area of technology for project development using various tools results in better modeling.
100003/ CS722U.1-PO2	M	Knowledge acquired in the selected area of project development can be used to identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions.
100003/ CS722U.1-PO3	M	Can use the acquired knowledge in designing solutions to complex problems.

100003/ CS722U.1-PO4	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1-PO5	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.1-PO6	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.1-PO7	M	Project development based on societal and environmental context solution identification is the need for sustainable development.
100003/ CS722U.1-PO8	L	Project development should be based on professional ethics and responsibilities.
100003/ CS722U.1-PO9	L	Project development using a systematic approach based on well defined principles will result in teamwork.
100003/ CS722U.1-PO10	M	Project brings technological changes in society.
100003/ CS722U.1-PO11	H	Acquiring knowledge for project development gathers skills in design, analysis, development and implementation of algorithms.

100003/ CS722U.1-PO12	H	Knowledge for project development contributes engineering skills in computing & information gatherings.
100003/ CS722U.2-PO1	H	Knowledge acquired for project development will also include systematic planning, developing, testing and implementation in computer science solutions in various domains.
100003/ CS722U.2-PO2	H	Project design and development using a systematic approach brings knowledge in mathematics and engineering fundamentals.
100003/ CS722U.2-PO3	H	Identifying, formulating and analyzing the project results in a systematic approach.
100003/ CS722U.2-PO5	H	Systematic approach is the tip for solving complex problems in various domains.
100003/ CS722U.2-PO6	H	Systematic approach in the technical and design aspects provide valid conclusions.
100003/ CS722U.2-PO7	H	Systematic approach in the technical and design aspects demonstrate the knowledge of sustainable development.
100003/ CS722U.2-PO8	M	Identification and justification of technical aspects of project development demonstrates the need for sustainable development.

100003/ CS722U.2-PO9	H	Apply professional ethics and responsibilities in engineering practice of development.
100003/ CS722U.2-PO11	H	Systematic approach also includes effective reporting and documentation which gives clear instructions.
100003/ CS722U.2-PO12	M	Project development using a systematic approach based on well defined principles will result in better teamwork.
100003/ CS722U.3-PO9	H	Project development as a team brings the ability to engage in independent and lifelong learning.
100003/ CS722U.3-PO10	H	Identification, formulation and justification in technical aspects will be based on acquiring skills in design and development of algorithms.
100003/ CS722U.3-PO11	H	Identification, formulation and justification in technical aspects provides the betterment of life in various domains.
100003/ CS722U.3-PO12	H	Students are able to interpret, improve and redefine technical aspects with mathematics, science and engineering fundamentals for the solutions of complex problems.

100003/ CS722U.4-PO5	H	Students are able to interpret, improve and redefine technical aspects with identification formulation and analysis of complex problems.
100003/ CS722U.4-PO8	H	Students are able to interpret, improve and redefine technical aspects to meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
100003/ CS722U.4-PO9	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.4-PO10	H	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools for better products.
100003/ CS722U.4-PO11	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.4-PO12	H	Students are able to interpret, improve and redefine technical aspects for demonstrating the knowledge of, and need for sustainable development.
100003/ CS722U.5-PO1	H	Students are able to interpret, improve and redefine technical aspects, apply ethical principles and commit to

		professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.5-PO2	M	Students are able to interpret, improve and redefine technical aspects, communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
100003/ CS722U.5-PO3	H	Students are able to interpret, improve and redefine technical aspects to demonstrate knowledge and understanding of the engineering and management principle in multidisciplinary environments.
100003/ CS722U.5-PO4	H	Students are able to interpret, improve and redefine technical aspects, recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
100003/ CS722U.5-PO5	M	Students are able to interpret, improve and redefine technical aspects in acquiring skills to design, analyze and develop algorithms and implement those using high-level programming languages.
100003/ CS722U.5-PO12	M	Students are able to interpret, improve and redefine technical aspects and contribute their engineering skills in computing and information engineering domains like

		network design and administration, database design and knowledge engineering.
100003/ CS722U.6-PO5	M	Students are able to interpret, improve and redefine technical aspects and develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life.
100003/ CS722U.6-PO8	H	Students will be able to associate with a team as an effective team player for the development of technical projects by applying the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
100003/ CS722U.6-PO9	H	Students will be able to associate with a team as an effective team player to Identify, formulate, review research literature, and analyze complex engineering problems
100003/ CS722U.6-PO10	M	Students will be able to associate with a team as an effective team player for designing solutions to complex engineering problems and design system components.
100003/ CS722U.6-PO11	M	Students will be able to associate with a team as an effective team player, use research-based knowledge and research methods including design of experiments, analysis and interpretation of data.

100003/ CS722U.6-PO12	H	Students will be able to associate with a team as an effective team player, applying ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.1-PSO1	H	Students are able to develop Computer Science Specific Skills by modeling and solving problems.
100003/ CS722U.2-PSO2	M	Developing products, processes or technologies for sustainable and socially relevant applications can promote Programming and Software Development Skills.
100003/ CS722U.3-PSO3	H	Working in a team can result in the effective development of Professional Skills.
100003/ CS722U.4-PSO3	H	Planning and scheduling can result in the effective development of Professional Skills.
100003/ CS722U.5-PSO1	H	Students are able to develop Computer Science Specific Skills by creating innovative solutions to problems.
100003/ CS722U.6-PSO3	H	Organizing and communicating technical and scientific findings can help in the effective development of Professional Skills.