PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

LARBI TEBESSI UNIVERSITY

FACULTY OF EXACT SCIENCES AND SCIENCES OF NATURE AND LIFE

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

TAKALEM : Deep learning approach for Algerian sign language recognition

Submitted by

Fathi Abdelmalek

Supervised by

Pr. Mohamed Amroun

Dr. Issam Bendib

Dr. M Yassine Haouam

Dr. Khelifa Boudjemaa

June 19, 2023

Dedication

To my dear mother

To my dear father

To my dear sisters and brothers

To my adorable grandmother

To all my uncles, aunts and cousins

To all my friends and colleagues

To every one who have supported me

FATHI ABDELMALEK

Acronyms

AJSL Algerian jewish sign language

ANN Artificial Neural Network

AR Augmented Reality

ASL American sign language

ASP Algerian sign language

CNN Convolutional Neural Network

CRF Conditional Random Field

DL Deep Learning

DMP Digital Motion Processor

DNN Deep Neural Network

FSL French sign language

HMM Hidden Markov Model

IMU Inertial Measurement Unit

IoT Internet of Things

LSTM Long-Short Term Memory

MEMS Micro Electro-Mechanical Systems

ML Machine Learning

NN Neural Network

RNN Recurrent Neural Network

SL Sign language

SLR Sign language recognition

SoC system-on-a-chip

SVM Support Vector Machines

VR Virtual Reality

Abstract

Sign language recognition (SLR) plays a crucial role in facilitating communication between individuals with hearing impairments and non-signers. However, the lack of an Algerian sign language (ASP) dataset and limited time constraints pose challenges in developing a customized dataset. In this study, an existing American sign language (ASL) dataset is utilized, and a Long-Short Term Memory (LSTM) model is employed for sign classification. The LSTM model achieves an impressive accuracy of 95.03% for character-level classification and 94.63% for word-level classification. Remarkably, the model exhibits minimal errors in both character and word classifications, primarily due to the selection of the most frequently predicted sign from a pool of 150 predictions per gesture. While these results demonstrate the effectiveness of the model.

Keywords— Sign Language Recognition, Deep Learning, LSTM

Contents

List of Figures

List of Tables

General Introduction

1 Context of Work

SLR is a rapidly evolving field that aims to enable effective communication between individuals with hearing impairments who use Sign language (SL) and those who do not. SL, including ASP, are visual languages that rely on hand gestures, facial expressions, and body movements to convey meaning. Recognizing and interpreting SL gestures accurately is essential for facilitating inclusive communication and ensuring equal access to information and services for the deaf community.

In recent years, advancements in Deep Learning (DL) have revolutionized SLR. DL models, particularly Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), have shown remarkable success in analyzing and understanding complex visual data, making them well-suited for SLR tasks. By training DL models on large SL datasets, researchers have achieved significant improvements in accuracy and real-time recognition.

However, the development of accurate and robust SLR systems, especially for less-studied SLs like ASP, remains a challenge. ASP has its unique vocabulary, grammar, and cultural aspects, which necessitate specific recognition models tailored to its characteristics. Additionally, limited resources, such as annotated ASP datasets, pose obstacles in building effective recognition systems.

To address these challenges, our research focuses on developing a DL model specifically designed for ASL recognition. Leveraging the advancements in DL and the availability of sensor-based ASL datasets, we aim to adapt and optimize these techniques for ASP recognition. By doing so, we contribute to bridging the communication gap and empowering the Algerian deaf community.

Furthermore, we aim to implement the developed DL model on the ESP32 microcontroller, which is widely used in embedded systems. By deploying the model on a low-power and portable device, we enable real-time SLR, making it accessible in various settings, including smartphones and wearable devices like smart gloves, and this is our objective. This implementation on microcontrollers opens up possibilities for ubiquitous and on-the-go SLR, further enhancing the integration and inclusion of the deaf community in everyday life.

In the following sections of this report, we present the state of the art in SLR, related works in the field, the conception of the TAKALEM gloves, the development of the DL model, and the evaluation of its performance. Through our research, we strive to contribute to the advancement of ASP recognition technology and promote inclusivity and accessibility for individuals with hearing impairments.

2 Problematic

The lack of existing approaches for SL recognition necessitates starting from scratch in the development of this systems. Furthermore, the limited availability of the needed materials to construct our own dataset and the time constraints we faced led us to leverage a sensor-based ASL dataset. Although this dataset differs from ASP, it serves as a foundation for our research and enables us to explore the effectiveness of DL techniques in interpreting and recognizing SL gestures.

3 Objectives

The main objectives of our research are as follows:

- Develop a DL model for ASL recognition.
- Investigate the effectiveness of DL techniques in interpreting and recognizing the gestures.
- Evaluate the performance of the developed model using appropriate metrics and validate its accuracy and robustness.
- Implement the developed model on the ESP32 microcontroller to enable real-time SLR.
- Contribute to the advancement of ASP recognition technology and facilitate effective communication between SL users and the wider community.

4 Structure of the Report

The report is structured as follows:

- **Part 1: State of the Art**
  - Chapter 1: Sign Languages and Sign Language Recognition.
  - Chapter 2: Related Works.
- **Part 2: Contribution**
  - Chapter 3: Conception of TAKALEM Gloves.

Part I

State of the art

Chapter 1

Sign languages and sign language recognition

1 Introduction

Deaf, mute, and hard-of-hearing people, have their language system, its SL, there are over 300 different SL used in the world, and each one has a unique grammar and vocabulary, one of them is ASP, is used by more than 240,000 individuals around Algeria "[5]. SLR technology aims to provide a means of automatically interpreting SL and translating it into text or speech, to facilitate the communication process between normal and deaf/mute people.

Recent advancements in wearable sensor technology have made it possible to develop SLR systems using sensor gloves. These gloves contain multiple sensors that measure various parameters related to hand movement and position. Machine Learning (ML) and DL algorithms can then be used to analyze the sensor data and recognize the corresponding SL gestures.

In this chapter, we will provide an overview of sign languages and SLR technology. We will discuss the challenges associated with SLR and the potential benefits of developing such technology.

2 Sign languages

SL is one of the methods of communication, which is defined as a set of visual symbols or gestures that are used in a very systematic way for words, concepts, or ideas of a language "[6].

Despite the complexity of sign languages, they can be broken down into smaller units, such as signs, hand shapes, and movements. Sign languages typically use a combination of these units to form words and sentences. For example, ASP uses hand shapes, movements, and facial expressions to convey meaning.

Sign languages are not just visual representations of spoken languages, they are unique and independent languages with their syntax, grammar, and vocabulary. Recognizing and understanding them is therefore crucial for effective communication between hearing and deaf communities. In recent years, there has been increasing interest in developing technology to aid sign language recognition and translation.

2.1 Most sign languages used in the world

Around 5% of the world population are deaf mute people "[4], they use different SLs as their communication languages. that make a variety of SLs used in the world. In table 1.1 we introduce the most used SLs in the world.

Table 1.1: Major sign languages of the world "[4]

Country	**Sign Language**	**Abbn**
United Kingdom	British Sign Language	BSL
United States of America	American Sign Language	ASL
Commonwealth of Australia	Australian Sign Language	Auslan
Japan	Japanese Sign Language	JSL
People's Republic of China	Chinese Sign Language	CSL
Taiwan	Taiwanese Sign Language	TSL
Middle-East	Arabic Sign Language	ArSL
Islamic Republic of Iran and other Gulf countries	Persian Sign Language	PSL
Republic of India	Indian Sign Language	ISL

2.2 Sign languages in Algeria

In Algeria, there is more than 240,000 deaf mute people "[5]. However there was no official SL for the country until 2002, when the Algerian government recognized the ASP as official SL in Algeria "[7]. But this language is not the only one the country, there is also Algerian jewish sign language (AJSL), an old SL which eveloped in several Jewish communities is the region of M'zab, Algeria, which is located in the northern part of the Sahara desert "[8].

For ASP, there is no official document or reference except for a dictionary published recently by the Algerian government. It contains some signs used by the deaf and other signs borrowed from the old French sign language (FSL) "[9]. This book consists of illustrations for gestures organized in categories like religion, justice, education, etc. If the gesture require hand movements, they support the illustration with arrows to show the direction of the movement.

2.3 Algerian Sign Language

ASP is an SL derived from FSL, used by the deaf/mute community of Algeria. It was officially recognized by the Algerian law as official SL in Algeria in May 2002 "[7]. Technically, it is a visual-gestural language that uses hand shapes, movements, and facial expressions to convey meaning. Therefore, this community is often excluded from basic communication, this has caused many deaf Algerians to go

without access to education, employment opportunities, and other basic rights. For that, The government of Algeria opened many deaf schools around the country to teach them the language itself, and basic education like any normal person.

Even with the existence of deaf schools, the teachers themselves are not qualified neither master ASP, all of them are hearing individuals who hold different degrees which are not related to deaf education or SL, and they have never been trained to use the language before they get hired, some teachers attend training courses to learn alphabet only. And this makes it difficult for those pupils to get basic education and go even to middle or high school and college, most of them are marginalized and can only be manual workers, they are denied access to a high-quality education that meets their special needs to improve their lives and live as equal to their peers "[9].

Figure 1.1: Algerian Sign Language "[1]

3 Sign language recognition

3.1 Definition

SLR is the process of interpreting and translating the gestures, movements, and facial expressions of SL into written or spoken language. It involves capturing, processing, and analyzing data from various sensors and devices such as gloves and cameras.

The task of SLR is a challenging one due to the complexity and variability of sign languages. They are rich and expressive, and there are many different sign languages used around the world, each with their own unique vocabulary, grammar, and syntax. Moreover, sign languages are not universal, meaning that a sign used in one language may have a completely different meaning in another language.

Despite these challenges, significant progress has been made in the field of SLR in recent years, thanks to advances in sensor technology, computer vision, ML and DL. Researchers have proposed a wide range of approaches to tackle the problem of SLR, including rule-based systems, template matching, Hidden Markov Model (HMM), Artificial Neural Network (ANN), and DL methods.

In recent years, DL-based methods, particularly CNN and RNN, have shown promising results in SLR, achieving state-of-the-art performance on several benchmark datasets. These methods can learn meaningful representations of the data directly from raw input, which makes them well-suited to complex and dynamic data like SL.

3.2 SLR applications and fields of use

SLR technology has the potential to empower deaf and hard-of-hearing individuals by enabling them to communicate more effectively with the wider community. Moreover, it can be used in other areas alongside basic use, here are the main areas of use of SLR technology.

3.2.1 SL translation

SL translation is one of the primary applications of SLR technology. By capturing and analyzing SL gestures, SLR systems can convert them into written or spoken language, enabling effective communication between deaf and hearing individuals. This technology can be utilized in various contexts, such as educational settings, customer service centers, healthcare facilities, and public institutions. For example, in education, SLR translation can support deaf students by providing real-time SL interpretation during lectures, ensuring they have access to the same educational content as their hearing peers. In customer service, SLR translation can facilitate communication between deaf customers and service representatives, improving accessibility and customer satisfaction. Additionally, SLR translation can be integrated into translation apps or devices, allowing deaf individuals to communicate with individuals who do not understand SL, bridging the communication gap.

3.2.2 Virtual and augmented reality

SLR technology can enhance the immersive experience of virtual and augmented reality environments by incorporating SLR capabilities. In VR/AR applications, SLR enables users to interact with the virtual world using SL gestures, making the experience more intuitive and inclusive for deaf users. For instance, in a Virtual Reality (VR) game, SLR can recognize and interpret SL commands as input, allowing players to control their characters or perform actions using SL gestures. In Augmented Reality (AR), SLR can be used to overlay real-time SL translations onto the user's field of view, enabling seamless communication between deaf and hearing individuals in AR scenarios. This integration of SLR in VR/AR not only enhances entertainment experiences but also opens up new possibilities in training simulations, remote collaboration, and interactive storytelling.

3.2.3 Research and linguistics

SLR technology plays a crucial role in linguistic research and the study of sign languages. By capturing and analyzing SL data, SLR systems provide valuable insights into the structure, grammar, and syntax of sign languages. Researchers can use SLR to examine the linguistic patterns and variations within SL communities, contributing to the documentation and preservation of sign languages. SLR can also assist in studying the cognitive aspects of SL processing and acquisition. Furthermore, SLR can be used to create SL corpora and databases, which serve as valuable resources for linguistic analysis and comparison across different SL systems. Overall, SLR technology empowers researchers and linguists to delve deeper into the intricate nature of sign languages, leading to a better understanding and appreciation of deaf culture and communication.

4 Deep Learning for SLR

DL, a subfield of ML, has revolutionized various domains by enabling the automatic learning of intricate patterns and representations from raw data. In SLR, DL proves to be a promising approach due to its ability to process and analyze complex visual and temporal information.

Neural Network (NN) serve as the foundation of DL models. These computational structures are inspired by the biological neurons in the human brain and consist of interconnected layers of artificial neurons. CNN are widely employed in image recognition tasks, including the analysis of hand gestures in SL. CNNs leverage convolutional layers to automatically learn spatial features from input images, enabling the model to discern intricate patterns and variations.

In addition to CNNs, RNN play a vital role in SLR. RNNs are particularly effective in handling sequential data, which is inherent in SL gestures. RNNs can capture the temporal dependencies in a sequence of gestures and retain information over time. LSTM networks, a type of RNN, excel in modeling long-range dependencies and have demonstrated promising results in SLR tasks.

Training and optimization are essential steps in the DL workflow. During the training phase, a NN is exposed to labeled training data, and its internal parameters are adjusted to minimize the discrepancy between predicted and actual outputs. Backpropagation, a gradient-based optimization algorithm, is commonly used to update the model's parameters. The training process is iterative, and the model undergoes multiple epochs to improve its performance gradually.

Despite the remarkable potential of DL in SLR, it comes with its own set of challenges and limitations. One major challenge is the scarcity of annotated SL datasets, which hinders the training process and necessitates domain-specific data collection efforts. Overfitting, a phenomenon where the model performs well on training data but fails to generalize to unseen data, is another challenge that requires careful regularization techniques.

Moreover, DL models often demand significant computational resources, including high-performance computing units and memory capacity. The interpretability of DL models is another concern, as they are often seen as "black boxes" that make it challenging to understand the internal decision-making process.

In conclusion, DL techniques offer immense potential for SLR, enabling the automatic extraction of meaningful representations from visual and temporal data. This section has provided a comprehensive overview of the key concepts, methodologies, and challenges associated with DL in the context of SLR. By leveraging DL, we can unlock new possibilities for accurate and real-time interpretation of SL gestures.

5 Conclusion

In this chapter, we have introduced sign languages and the importance of SLR systems in facilitating communication between deaf or hard-of-hearing individuals and the hearing world. We have also presented a review of the literature on SLR systems, including the techniques and methodologies that have been used to develop these systems.

From our review, it is clear that SLR is a challenging task that requires the use of sophisticated techniques such as ML, computer vision, and signal processing. While progress has been made in this area, there are still many open challenges that need to be addressed, such as improving the accuracy and robustness of recognition systems, developing systems that can recognize different sign languages, and addressing the issue of data sparsity.

In summary, the field of SLR is a promising area of research with many potential applications. We hope that this chapter has provided the reader with a good understanding of the current state of the art in SLR and the challenges that lie ahead.

Chapter 2

Related works

1 Introduction

Sign language recognition refers to the process of interpreting and understanding sign language gestures to facilitate communication between individuals who are deaf or hard of hearing and those who are not proficient in sign language. Sign languages are complex visual languages with distinct grammar, syntax, and cultural variations. Recognizing and translating sign language gestures in real-time present significant challenges due to the intricacies and nuances involved.

The development of robust and accurate sign language recognition systems has gained increasing attention in recent years due to the potential impact on inclusive education, healthcare accessibility, smart home integration, and emergency communication. By bridging the communication gap between signers and non-signers, these systems contribute to creating an inclusive environment for individuals with hearing impairments.

1.1 Motivation for Exploring Related Works

Exploring related works in sign language recognition is essential for several reasons. Firstly, understanding the existing literature and research provides valuable insights into the state of the field, including advancements, methodologies, and limitations. This knowledge enables researchers to build upon previous findings and avoid duplicating efforts.

Secondly, analyzing related works helps identify gaps and challenges in current approaches to sign language recognition. By understanding the limitations of existing methods, researchers can propose innovative solutions to address these limitations and enhance the overall accuracy, speed, and usability of recognition systems.

Furthermore, advancements in sign language recognition technology have the potential to transform various sectors. In education, sign language recognition systems can assist in teaching sign language to non-signers, facilitating inclusive classrooms and promoting sign language literacy. In healthcare settings,

these systems enable effective communication between healthcare providers and patients with hearing impairments, enhancing the quality of care and patient outcomes.

Moreover, integrating sign language recognition into smart homes and devices enhances accessibility and convenience for individuals who rely on sign language as their primary mode of communication. Additionally, in emergency situations where verbal communication may be challenging or impossible, sign language recognition systems can play a crucial role in ensuring effective communication and timely assistance.

1.2 Objectives of the Chapter

This chapter aims to provide a comprehensive overview of sign language recognition by exploring related works in the field. The specific objectives are as follows:

1.2.1 Explore Datasets for Training and Evaluation

The chapter will investigate existing sign language datasets utilized for training and evaluating sign language recognition systems. It will discuss the characteristics of representative datasets, such as size, diversity, annotation methods, and challenges associated with dataset collection and annotation.

1.2.2 Review and Analyze Existing Methodologies

The chapter will examine various methodologies and techniques used in sign language recognition, including vision-based methods, data glove-based methods, and hybrid approaches. By analyzing these methodologies, their strengths, limitations, and performance, researchers can gain insights into the different approaches used in the field.

2 Datasets for Training and Evaluation

In the context of SLR, synthesis refers to the process of generating natural and expressive SL gestures from textual or spoken input. Synthesis plays a vital role in bridging the communication gap between SL users and non-signers by providing a means for converting spoken or written language into SL.

2.1 Text-to-Sign Synthesis Methods

Text-to-sign synthesis methods focus on converting written text into SL gestures. These methods involve linguistic analysis of the input text to determine the appropriate sign vocabulary and grammatical structure. Various techniques, such as rule-based systems, statistical models, and ML algorithms, have been employed in text-to-sign synthesis.

Rule-based systems utilize predefined rules and linguistic knowledge to map words or phrases to corresponding signs. These systems typically rely on manually crafted linguistic resources, such as sign dictionaries and grammatical rules, to generate SL sequences.

Statistical models, such as HMM and Conditional Random Field (CRF), have been used to learn the statistical relationships between textual input and SL output. These models are trained on annotated corpora of text-sign pairs, allowing them to capture the patterns and dependencies between words and signs.

ML algorithms, including ANN and DL models, have shown promising results in text-to-sign synthesis. These models can learn the mapping between text and SL directly from data, leveraging large-scale annotated datasets to improve the quality and naturalness of generated SL gestures.

2.2 Speech-to-Sign Synthesis Methods

Speech-to-sign synthesis methods aim to convert spoken language into SL gestures. These methods involve audio analysis and processing techniques to extract relevant features from the input speech signals. The extracted features are then mapped to corresponding SL gestures using statistical models or ML algorithms.

Speech recognition algorithms, such as hidden HMMs and Deep Neural Network (DNN)s, are often used to transcribe spoken language into textual representations. The resulting text can then be processed using text-to-sign synthesis methods to generate SL gestures.

Multimodal approaches, combining audio and visual information, have also been explored in speech-to-sign synthesis. These approaches leverage audiovisual recordings of SL performances to learn the mapping between speech and SL gestures. By incorporating both acoustic and visual features, these methods can capture the nuances of spoken language and produce more accurate and natural SL outputs.

3 Methods of sign language recognition

SLR methods encompass a range of techniques and approaches used to interpret and understand SL gestures. These methods can be categorized into several broad categories, including vision-based methods, data glove-based methods, and hybrid approaches that combine multiple modalities for improved accuracy and robustness.

3.1 Vision-Based Sign Language Recognition Methods

Vision-based SLR methods rely on computer vision techniques to analyze and interpret SL gestures captured by video cameras or depth sensors. These methods typically involve extracting relevant features from the visual data and mapping them to corresponding signs in a predefined SL vocabulary.

Various techniques have been employed in vision-based approaches, including hand shape analysis, motion analysis, and spatiotemporal modeling. Hand shape analysis focuses on extracting information related to hand shape and configuration, such as fingertip positions, hand contours, and hand landmarks. Motion analysis techniques capture the dynamics of hand movements, including trajectory, speed, and acceleration. Spatiotemporal modeling methods aim to capture the spatial and temporal relationships between different hand movements and gestures.

3.2 Data Glove-Based Sign Language Recognition Methods

Data glove-based SLR methods involve the use of sensor-equipped gloves to capture and analyze hand movements during SL production. These gloves are equipped with sensors, such as flex sensors or Inertial Measurement Unit (IMU), which measure the bending angles of fingers or capture hand orientation and motion.

Flex sensor-based methods utilize the bending angles of individual fingers to recognize SL gestures. These sensors provide information about the flexion and extension of each finger, which can be used to distinguish different signs based on finger configurations.

IMU-based methods utilize inertial sensors, such as accelerometers and gyroscopes, to capture hand motion and orientation. These sensors provide data on the acceleration, angular velocity, and orientation of the hand, which can be used to infer SL gestures.

3.3 Hybrid Approaches in Sign Language Recognition

Hybrid approaches combine multiple modalities, such as vision and data glove sensors, to enhance the accuracy and robustness of SLR systems. By integrating visual information with data from sensors, these methods can capture both fine-grained hand movements and global hand positions, leading to more comprehensive representations of SL gestures.

Hybrid approaches often involve fusing the data from different modalities at different levels, such as feature fusion, decision-level fusion, or early fusion. Feature fusion combines the extracted features from vision and sensor data to create a unified feature representation. Decision-level fusion combines the decisions made by individual classifiers trained on different modalities to make a final decision. Early fusion combines the raw data from different modalities at the input level to create a joint representation for further processing.

4 Synthesis

Synthesis in SLR refers to the process of combining different approaches and techniques to achieve better results. With the increasing interest in DL, researchers have incorporated ANN models such as LSTM into SLR systems.

In addition to using LSTM, researchers also employ other methods such as HMM, Support Vector Machines (SVM), and CNN. By combining these various techniques, it is possible to improve accuracy rates and reduce error rates in sign language recognition.

Moreover, datasets play a crucial role in synthesis since they provide a means for training and testing models. Researchers use publicly available datasets or create their own annotated corpus of data for specific purposes.

Evaluation metrics are used to measure the performance of synthesized systems. Metrics like precision, recall and F1-score help researchers assess the effectiveness of their approach.

The development of synthesized systems has great potential for improving communication between hearing-impaired individuals and those who do not know SL. Future research studies on this topic will continue exploring new ways of integrating different techniques that enable accurate recognition algorithms with faster processing times than current state-of-the-art solutions can offer.

5 Conclusion

SLR is a complex task that involves the use of advanced algorithms and techniques in computer vision, ML, and DL. The development of SLR systems has made it possible for the deaf community to communicate more effectively with people who do not understand SL.

We have discussed the different methods used in SLR such as ANN and LSTM. We have also talked about some datasets commonly used in SLR research, as well as evaluation metrics that are frequently employed to assess the performance of these systems.

There have been significant advancements in DL-based approaches for SLR over recent years. However, much work still needs to be done to improve accuracy rates further. More comprehensive data sets need to be created with more diverse sets of gestures so that models can learn better.

All things considered; we believe that continued research into new algorithms and techniques will pave the way for even more sophisticated SLR systems capable of discerning subtle nuances between signs accurately.

Part II

Contribution

Chapter 3

Conception of TAKALEM gloves

1 Introduction

SLs play a crucial role in facilitating communication for individuals with hearing impairments. However, comprehending and interpreting SL poses challenges for those who are not familiar with its intricacies. To address this, there is a growing demand for technology-based solutions that can enable SLR and enhance communication between signers and non-signers.

The primary objective of this chapter is to introduce the conception of TAKALEM Gloves, a wearable technology-based system for SLR. These gloves are designed to capture hand gestures and translate them into corresponding signs or text in real-time, enabling seamless communication between individuals who use SL and those who do not.

This chapter focuses on the hardware architecture and configuration of the TAKALEM Gloves. It explores the technological aspects of the gloves, including the selection of microcontrollers, sensors, and other essential components. Understanding the underlying hardware is crucial for gaining insights into the system's capabilities and limitations.

Furthermore, the motivation behind the development of TAKALEM Gloves is explored in this chapter. The ability to bridge the communication gap between signers and non-signers holds significant implications in various domains. In the field of education, the gloves can be utilized to teach SL to non-signers, fostering inclusivity and promoting accessibility in educational settings. Within the healthcare sector, the gloves facilitate effective communication between healthcare providers and patients with hearing impairments, enhancing the quality of care and patient experience.

Moreover, the integration of TAKALEM Gloves into smart homes and devices offers opportunities for seamless interaction and control. Users can communicate with their smart devices using SL, enhancing convenience and accessibility in the context of home automation. Additionally, the gloves prove invaluable

in emergency situations where communication is critical, facilitating swift and efficient communication between individuals in high-stress scenarios.

This chapter sets the stage for the subsequent sections, which will delve into specific aspects of the SLR system. By understanding the significance of SLs, the diverse applications of TAKALEM Gloves, and the underlying motivation, we can recognize the importance of developing an accurate and reliable system that empowers individuals with hearing impairments.

2 Hardware architecture and configuration

The TAKALEM Gloves, designed for SLR, consist of two main units: the sensing unit and the processing unit. This section focuses on the hardware architecture and configuration of these units, providing an overview of the components and their interconnections.

The sensing unit of the TAKALEM Gloves is responsible for capturing and collecting data related to hand movements and gestures. It comprises flex sensors and an MPU6050 sensor. The flex sensors are strategically positioned on the glove to measure the bend of individual fingers. They are connected to the microcontroller through designated pins, namely pins 13, 15, 25, 26, and 27. These flex sensors provide precise and real-time measurements of finger movements, allowing for accurate interpretation of SL gestures.

In addition to the flex sensors, the TAKALEM Gloves also incorporate the MPU6050 sensor. The MPU6050 is an IMU that combines a three-axis accelerometer and a three-axis gyroscope. It measures the orientation and acceleration of the hand, providing valuable data for the SLR process. This sensor is connected to the microcontroller via the ASL (Accelerometer Serial Interface) and ASD (Accelerometer Serial Data) pins, enabling seamless data communication between the sensor and the processing unit.

The processing unit of the TAKALEM Gloves is powered by the ESP32 WROOM Module, a powerful microcontroller that integrates Wi-Fi and Bluetooth functionalities. This module acts as the brain of the gloves, processing the data collected from the flex sensors and the MPU6050 sensor. It is responsible for extracting and analyzing the sensor data, applying signal processing techniques, and calling the DL model to predict the corresponding sign.

The ESP32 WROOM Module is equipped with sufficient computational power and memory to handle the complex algorithms and computations involved in real-time SLR. Its connectivity features allow for seamless integration with other devices, such as smartphones or smart home systems, enabling a wide range of applications.

To ensure optimal performance and reliable data transmission, the hardware components are carefully configured and calibrated. The flex sensors are positioned and secured on the glove to accurately

measure finger bend without hindering natural hand movements. The MPU6050 sensor is properly calibrated to provide accurate orientation and acceleration data.

As a visual representation of the TAKALEM Gloves, 3.1 showcases the prototype of the gloves, highlighting the positioning of the flex sensors and the MPU6050 sensor. This prototype serves as a physical embodiment of the hardware architecture described in this section.

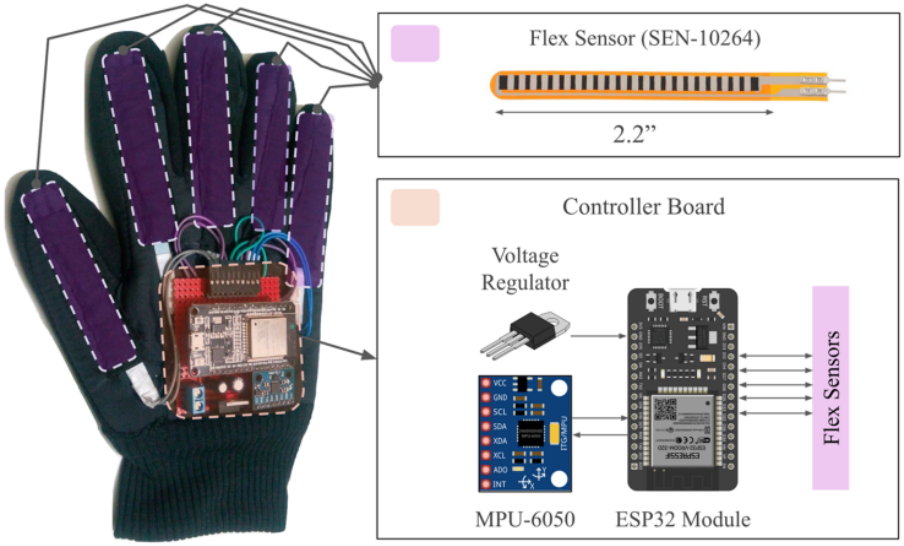


Figure 3.1: TAKALEM Gloves components architecture

2.1 ESP32 WROOM

ESP32 is a highly versatile and powerful microcontroller system-on-a-chip (SoC) developed by Espressif Systems. It combines Wi-Fi and Bluetooth connectivity with a dual-core processor, making it ideal for various Internet of Things (IoT) applications.

2.2 Flex sensor

Flex sensor is basically a variable resistor whose terminal resistance increases when the sensor is bent. So this sensor resistance increases depends on surface linearity. So it is usually used to sense the changes in linearity ”[2].

How flex sensor work As shown in figure 3.4, when the surface of the flex sensor is completely linear it will be having its nominal resistance. When it is bent 45^{o} angle the flex sensor resistance increases to twice as before. And when the bent is 90^{o} the resistance could go as high as four times the nominal resistance. So the resistance across the terminals rises linearly with bent angle. So in a sense the flex sensor converts flex angle to resistance parameter ”[2].

Flex sensor pinout The flex sensor is a two-way variable resistor with two pins, one for GND and the other for VCC. We wire one pin directly to the GND, the second pin is used to read the flex resistance

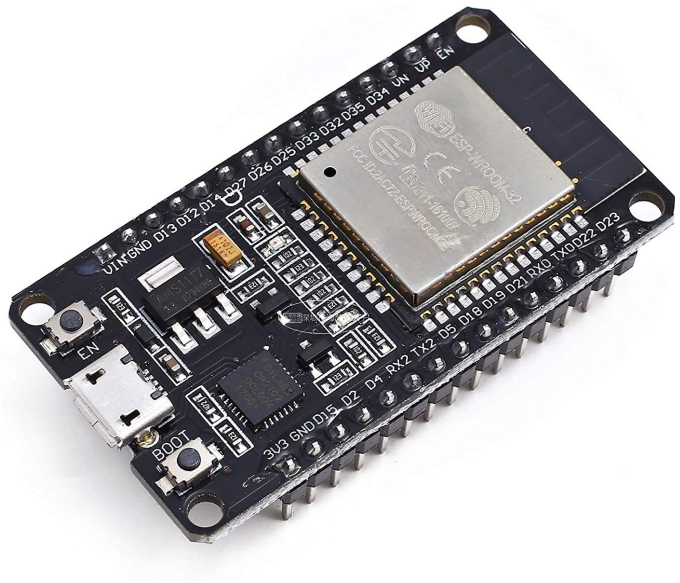


Figure 3.2: ESP32 WROOM

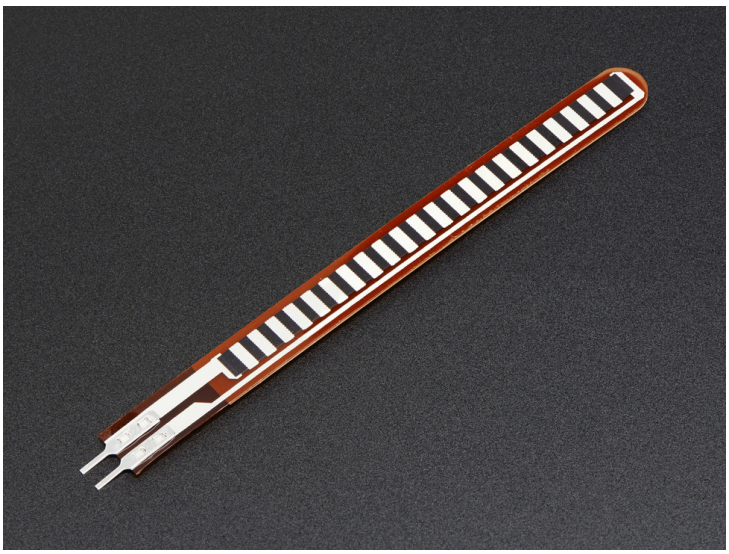


Figure 3.3: Flex sensor

via two wires, one is connected to the output pin in the esp32 and the other is connected to a constant resistance connected to the VCC.

2.3 MPU6050

The MPU6050 is a Micro Electro-Mechanical Systems (MEMS) which consists of a 3-axis Accelerometer and 3-axis Gyroscope inside it. This helps us to measure acceleration, velocity, orientation, displacement and many other motion related parameter of a system or object. This module also has a Digital Motion Processor (DMP) inside it which is powerful enough to perform complex calculation and thus free up the work for microcontrollers "[10].

Figure 3.4: Flex sensor cases "[2]

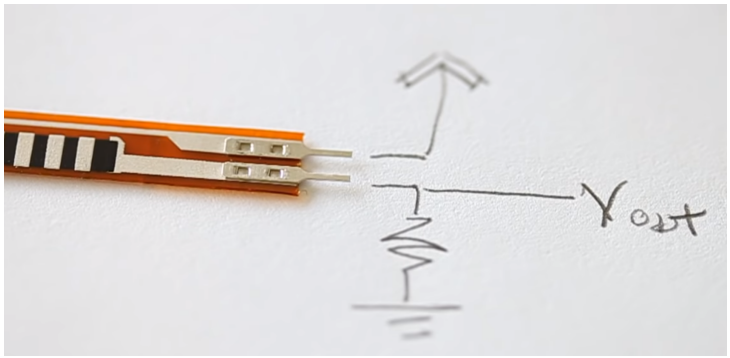


Figure 3.5: Flex sensor pin out

MPU6050 pinout The MPU6050 sensor module and ESP32 have specific pinouts that need to be connected correctly. The MPU6050 typically has four important pins: VCC, GND, SDA, and SCL. The VCC pin is connected to a 3.3V or 5V output on the ESP32 to power the sensor. The GND pin of the MPU6050 is connected to a ground pin on the ESP32 to establish a common ground. The SDA pin of the MPU6050 is connected to the SDA pin on the ESP32 (pin D23), which handles the data line for I2C communication. Similarly, the SCL pin of the MPU6050 is connected to the SCL pin on the ESP32 (pin D21), which handles the clock line for I2C communication. It's important to make sure the correct pins are connected to establish proper communication between the MPU6050 and ESP32.

3 Dataset collection and preprocessing

In this section, we provide an overview of the dataset used for training and evaluating the SLR system implemented with the TAKALEM Gloves. Due to time constraints and limited resources, we did not construct our own "ASP dataset." Instead, the dataset used in this project is the "ASL-Sensor-Dataglove Dataset", which is publicly available for use. The dataset consists of sensor readings captured during the performance of ASL gestures.

The "ASL-Sensor-Dataglove" Dataset contains a comprehensive collection of SL gestures, including 40 different signs derived from the ASL dictionary, it includes 26 letters of the alphabet and 14 commonly used words in SL. The dataset was collected from a group of 25 subjects, comprising 19 males and 6 females. Each subject performed each sign gesture a total of 10 times, resulting in a dataset of 10,000 records. Therefore, there are 250 records available for each gesture, considering the repetitions from each

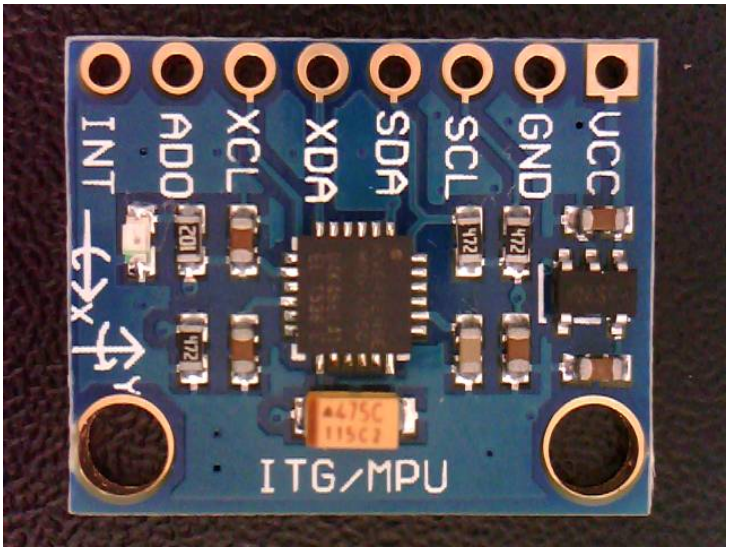


Figure 3.6: MPU6050

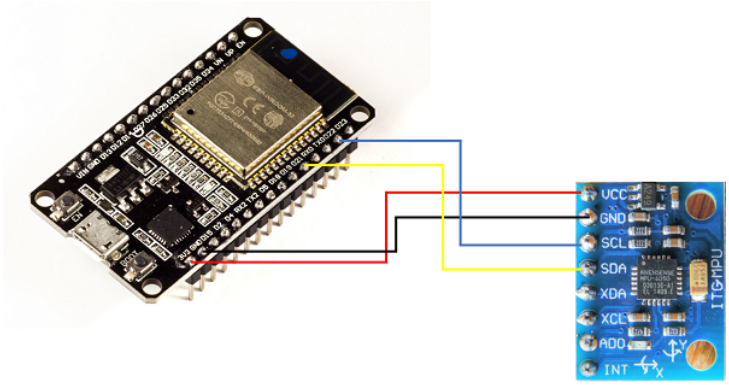


Figure 3.7: MPU6050 pin out "[3]

subject. The dataset was recorded using a dataglove equipped with various sensors, capturing important hand movements and finger positions during the performance of each gesture.

To ensure consistency and usability, each gesture recording has a duration of one second and a half. Therefore, each gesture is represented by 150 consecutive rows in the CSV file, and each CSV file contains 1500 rows, representing a total of 1,500,000 rows in the entire dataset. This large-scale dataset enables robust training and evaluation of sign language recognition models.

The dataset is made publicly available for researchers and practitioners in the field of sign language recognition. It can be accessed through the following link: "ASL-Sensor-Dataglove" Dataset. The dataset owners have graciously made it accessible for public use, facilitating further advancements in sign language recognition research and development.

Additionally, the dataset contained additional columns for timestamp and user identification. These columns helped in tracking the temporal aspects of the gestures and associating them with the respective subjects. However, for the purposes of our DL model, we primarily focused on the sensor data columns

Table 3.1: Summary of the ASL-Sensor-Dataglove Dataset

Dataset Size	10,000 records
Number of Signs	40
Subjects	25 (19 males, 6 females)
Records per Sign	250

Table 3.2: Selected Columns in the "ASL-Sensor-Dataglove" Dataset

Sensor Type	**Description**	**Number of Columns**
Flex Sensors	Measures finger bend	5
Gyro	Measures orientation and angular velocity	3
Acceleration	Measures acceleration along each axis	3

mentioned above.

Overall, the "ASL-Sensor-Dataglove" Dataset serves as a valuable resource for SLR research. Its availability to the public encourages collaboration and promotes the development of advanced models and algorithms in the field. The dataset's comprehensive nature, with 40 different signs and extensive recordings, allows for in-depth analysis and robust evaluation of SLR systems.

4 Proposed deep learning architecture

In this section, we present the DL architecture used for SLR using the TAKALEM Gloves. The architecture consists of an input layer, an LSTM layer, and an output layer, and it is designed for both words and characters classification.

Input Layer: The input layer receives the sensor data collected from the TAKALEM Gloves. The dataset is reshaped to have a shape of (3360, 150, 11), where 3360 represents the number of gestures used for training, 150 represents the number of rows for each gesture, and 11 represents the number of features. The features include the flex sensor readings, gyro data, and raw acceleration values.

LSTM Layer: The LSTM layer is a crucial component of the architecture as it captures the sequential information and temporal dependencies present in the SL gestures. It is responsible for learning the patterns and dynamics of hand movements and finger positions. The LSTM layer processes the input sequence and retains information over extended time periods, mitigating the vanishing gradient problem. In our model we used 38 nodes for this layer, we tried different number of nodes, and we noticed that the accuracy is getting higher as the number of nodes is getting higher, but we chosen exactly 38 because it's the highest possible number that the esp32 can handle duo to it's limits of flash memory.

Output Layer: The output layer is the final layer of the architecture and is responsible for generating predictions. For the words classification model, the output layer has 14 nodes, representing the 14 different words in the dataset. For the characters classification model, the output layer has 26 nodes, representing the 26 letters of the alphabet. The output layer applies a *softmax* activation function to

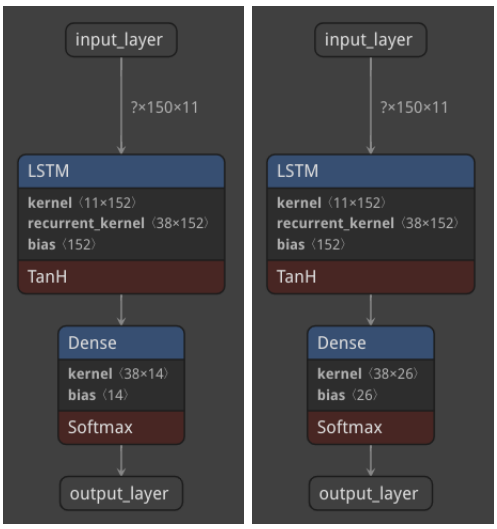
produce a probability distribution over the classes, with the highest probability indicating the predicted sign.

The proposed architecture is trained using the reshaped dataset, with appropriate training, validation, and testing splits. The model's parameters are adjusted during the training process using Adam as optimization algorithm, and "categorical cross-entropy" is chosen for loss function.

The performance of the architecture is evaluated using various evaluation metrics, including accuracy, precision, recall. These metrics assess the model's ability to correctly classify SL gestures and provide insights into its overall effectiveness.

The architecture is implemented using TensorFlow python library, which provide efficient tools for constructing and training ANNs like LSTM. The training process can benefit from powerful computing resources, such as GPUs, to accelerate the training time and handle the large-scale nature of the dataset .

By utilizing the proposed deep learning architecture in figure 3.8, the TAKALEM Gloves demonstrate promising results in SLR. The model's ability to accurately classify both words and characters is a significant step towards enabling effective communication for individuals using SL.



(a) Words model (b) Characters model

Figure 3.8: The trained model

In order to be able to use the model on the esp32, we should convert it to tensorflow lite version, and then convert it to a c++ array, as shown in figure 3.9, the tflite use different naming for different layers, the **LSTM** layer become **UnidirectionalSequenceLSTM**, and the **Dense** layer become **FullyConnected**. The size of the c++ array for the model is 37452 for characters model, and 35580 for words model.

(a) Words model (b) Characters model

Figure 3.9: The trained model after converted to tflite model

5 Evaluation criteria

In this section, we discuss the evaluation criteria used to assess the performance and effectiveness of the TAKALEM Gloves system for SLR. Evaluating the system's performance is crucial to understanding its strengths, limitations, and areas for improvement.

5.1 Accuracy

Accuracy is a fundamental metric used to evaluate the performance of SLR systems. It measures the system's ability to correctly classify and interpret SL gestures. In the context of the TAKALEM Gloves system, accuracy refers to the percentage of correctly recognized signs out of the total number of signs in the dataset. It can be calculated as in Eq. 3.1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Where:

- TP represents true positive, the number of correctly recognized signs.
- TN represents true negative, the number of correctly rejected non-target signs.
- FP represents false positive, the number of incorrectly recognized signs.
- FN represents false negative, the number of incorrectly rejected target signs.

5.2 Precision and macro-average precision

Precision refers to the measure of how accurate a model's predictions are, specifically the proportion of true positive predictions out of all positive predictions made by the model. It focuses on the correctness

of the predicted positive instances. Macro-average precision, on the other hand, is a way to calculate the average precision across multiple classes or categories in a classification problem. It involves calculating precision for each class independently and then averaging them without considering the class imbalance. Macro-average precision gives equal weight to each class, regardless of their size, providing a balanced evaluation metric for multi-class problems. They can be calculated as in Eq. 3.2 and Eq. 3.3:

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{3.2}$$

$$Precision_{macro-avg} = \frac{1}{N}\sum_{i=0}^{N} Precision_i \tag{3.3}$$

5.3 Recall and macro-average recall

Recall, also known as sensitivity or true positive rate, is a measure of how effectively a model identifies all relevant instances in a dataset. It calculates the proportion of true positive predictions out of all actual positive instances. Recall focuses on the ability of the model to correctly identify positive instances and avoid false negatives. Macro-average recall, similar to macro-average precision, is a method to compute the average recall across multiple classes or categories in a classification problem. It involves calculating recall for each class independently and then averaging them without considering the class imbalance. Macro-average recall provides an unbiased evaluation metric by giving equal weight to each class, regardless of their size or prevalence in the dataset. They can be calculated as in Eq. 3.4 and Eq. 3.5:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{3.4}$$

$$Recall_{macro-avg} = \frac{1}{N}\sum_{i=0}^{N} Recall_i \tag{3.5}$$

5.4 F1 score and macro-average f1 score

F1 score is a single metric that combines precision and recall into a single value, providing a balanced evaluation of a model's performance. It is the harmonic mean of precision and recall and is often used in situations where both precision and recall are important. Macro-average F1 score is a way to calculate the average F1 score across multiple classes or categories in a classification problem. It involves calculating the F1 score for each class independently and then averaging them without considering the class imbalance. Similar to macro-average precision and recall, macro-average F1 score gives equal weight to each class, providing a balanced evaluation metric that accounts for the performance across all classes, regardless of their size or prevalence in the dataset. They can be calculated as in Eq. 3.6 and Eq. 3.7:

$$F1_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{3.6}$$

$$F1_{macro-avg} = \frac{1}{N}\sum_{i=0}^{N} F1_i \tag{3.7}$$

By employing these evaluation criteria, including accuracy, precision, recall, F1 score, we can comprehensively assess the performance and capabilities of the TAKALEM Gloves system for SLR. The combination of these metrics enables us to gain a holistic understanding of the system's effectiveness, identify areas for improvement, and guide future enhancements in the field of SLR technology.

6 Results

In this section, we present the evaluation results for the word and character recognition models. The purpose is to assess the performance of each model and gain insights into their effectiveness.

6.1 Word Recognition Results

The word recognition model was evaluated using a comprehensive set of word samples. The model achieved an accuracy of 94.63% on the test dataset. The confusion matrix, depicted in Figure 3.10a, shows the distribution of predicted labels versus the true labels for each word class.

From the confusion matrix, we can observe that the model performs remarkably well, with most of the samples being correctly classified. There are only slight errors in some cases. However, overall, the model demonstrates high accuracy and precision.

To gain a comprehensive understanding of the word recognition model's performance, we present the classification report in Figure 3.11a.

The classification report provides insights into the precision, recall, and F1 score for each word class. The macro average precision, recall, and F1 score for the word recognition model are 94%, 95%, and 95%, respectively. These results indicate the model's strong performance in correctly recognizing the majority of the word classes. 3.11a.

6.2 Character Recognition Results

Moving on to character recognition, the model achieved an accuracy of 95.03% on the test dataset. The confusion matrix, shown in Figure 3.10b, illustrates the predicted labels versus the true labels for each character class.

Similar to the word recognition model, the confusion matrix for the character recognition model demonstrates excellent performance, with minimal errors. The model accurately recognizes the majority of the characters, with only slight confusion between similar characters like "u" and "v".

To gain a comprehensive understanding of the character recognition model's performance, we present the classification report in Figure 3.11b.

The classification report presents the precision, recall, and F1 score for each character class. The average precision, recall, and F1 score for the character recognition model are 95% for all of them. These results indicate the model's high accuracy in correctly identifying the majority of the character classes.

In summary, both the word and character recognition models demonstrate excellent performance with an accuracy of 94.63% and 95.03% respectively. The confusion matrices reveal minimal errors, while the classification reports confirm the models' precision, recall, and F1 score across various word and character classes. These results indicate the effectiveness and reliability of the implemented models in SLR.

7 Conclusion

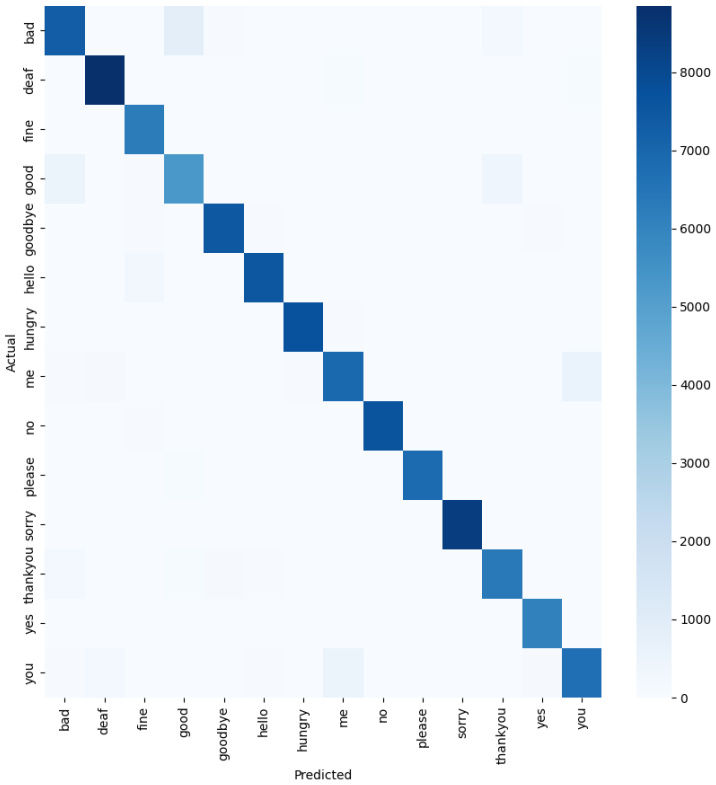
Our study aimed to develop a SLR model using a pre-existing sensor-based ASL dataset. Despite the time constraints and limited resources, we leveraged the ASL dataset, which consists of 40 different signs, including 26 letters and 14 common words. The dataset was collected from 25 subjects, with each subject repeating each gesture 10 times, resulting in a total of 150,000 records.

Throughout our research, we successfully trained and evaluated a DL model for SLR. The word recognition model achieved an impressive accuracy of 94.63% on the test dataset, while the character recognition model achieved an accuracy of 95.03%. These results demonstrate the effectiveness and reliability of our implemented models in accurately recognizing and classifying SL gestures.

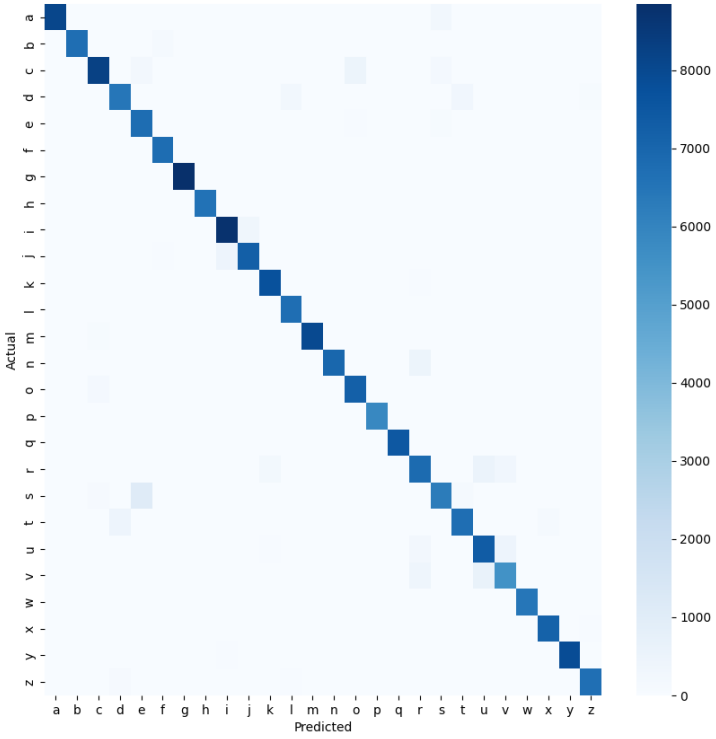
However, it is important to acknowledge the limitations of our study. The use of an ASL dataset, instead of a dedicated ASP dataset, poses a significant limitation. While the ASL dataset provided a solid foundation for our research, it may not fully capture the unique characteristics and nuances of ASP. Future work should prioritize the collection of a dedicated dataset specific to ASP to train the model more effectively.

Furthermore, considering the limited number of subjects and gestures in our dataset, expanding the dataset with more diverse subjects and a broader range of signs would contribute to a more comprehensive and generalized SLR model.

In conclusion, our study demonstrates the potential of DL models in SLR. Despite the limitations, our models achieved high accuracy in recognizing both words and characters from the ASL dataset. By addressing the identified limitations and incorporating additional data sources, such as dedicated ASP datasets and multiple modalities, future research can further improve the accuracy and applicability of SLR systems.



(a) Confusion matrix for the word recognition model.



(b) Confusion matrix for character recognition model

Figure 3.10: Confusion matrices

	precision	recall	f1-score	support
0	0.89	0.86	0.87	8550
1	0.96	0.98	0.97	9000
2	0.93	0.99	0.96	6300
3	0.82	0.84	0.83	6300
4	0.98	0.98	0.98	7650
5	0.98	0.96	0.97	7800
6	0.99	0.99	0.99	7800
7	0.91	0.89	0.90	7800
8	1.00	0.99	1.00	7650
9	0.99	0.97	0.98	7050
10	0.99	1.00	1.00	8400
11	0.90	0.93	0.91	6900
12	0.97	0.99	0.98	6150
13	0.91	0.87	0.89	7650
accuracy			0.95	105000
macro avg	0.94	0.95	0.95	105000
weighted avg	0.95	0.95	0.95	105000

(a) Classification report for word recognition model

	precision	recall	f1-score	support
0	1.00	0.96	0.98	8400
1	1.00	0.98	0.99	6900
2	0.95	0.89	0.92	9300
3	0.91	0.89	0.90	7200
4	0.83	0.98	0.90	6900
5	0.97	1.00	0.99	6750
6	1.00	1.00	1.00	8850
7	1.00	1.00	1.00	6600
8	0.94	0.96	0.95	9150
9	0.95	0.93	0.94	7800
10	0.96	0.99	0.97	7800
11	0.95	1.00	0.97	6750
12	1.00	0.99	0.99	8100
13	0.99	0.93	0.96	7500
14	0.93	0.96	0.94	7500
15	1.00	1.00	1.00	5850
16	1.00	1.00	1.00	7500
17	0.85	0.86	0.85	7950
18	0.91	0.82	0.86	7650
19	0.93	0.91	0.92	7350
20	0.86	0.91	0.88	8100
21	0.87	0.83	0.85	6600
22	1.00	1.00	1.00	6450
23	0.97	0.99	0.98	7200
24	0.99	0.99	0.99	7950
25	0.98	0.97	0.97	6900
accuracy			0.95	195000
macro avg	0.95	0.95	0.95	195000
weighted avg	0.95	0.95	0.95	195000

(b) Classification report for character recognition model

Figure 3.11: Classification reports

General Conclusion

In this report, we have tackled the challenge of SLR using DL techniques. Despite the lack of existing approaches for ASP recognition, we have successfully developed and evaluated a DL model using sensor-based ASL datasets. Our results demonstrate the potential of DL, particularly LSTM, in accurately recognizing ASL gestures. Moreover, we have implemented the model on the ESP32 microcontroller, enabling real-time SLR on low-power and portable devices. While the absence of a dedicated ASP dataset remains a limitation, our work serves as a foundation for future research in this domain. Collecting a comprehensive dataset and refining the recognition model for ASP are crucial next steps. By promoting inclusivity and accessibility, we aim to inspire further advancements in SLR and empower individuals with hearing impairments.

References

[1] "African sign lnaguages resource center - algeria." https://africansignlanguagesresourcecenter.com/profiles-of-african-sign-languages-and-deaf-culture-54algeria/.

[2] "Flex-sensor-components101." https://components101.com/sensors/flex-sensor-working-circuit-datasheet.

[3] "Interfacing esp32 with mpu6050." https://www.tutorialspoint.com/esp32_for_iot/interfacing_esp32_with_mpu6050.htm.

[4] A. Sahoo, G. Mishra, and K. Ravulakollu, "Sign language recognition: State of the art," *ARPN Journal of Engineering and Applied Sciences*, vol. 9, 02 2014.

[5] "Algerian sign language - ethnologue." https://www.ethnologue.com/25/language/asp/.

[6] H. Abdelouafi and M. Omari, "Teaching sign language to the deaf children in adrar, algeria."

[7] "Algerian sign language - wikipedia." https://en.wikipedia.org/wiki/Algerian_Sign_Language.

[8] S. Lanesman, *Algerian Jewish Sign Language: its emergence and survival*.

[9] H. Abdelouafi, "Challenges of deaf education in algeria."

[10] "Mpu6050-components101." https://components101.com/sensors/mpu6050-module.