**POPULAR AND DEMOCRATIC REPUBLIC OF ALGERIA**
**MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH**
**UNIVERSITY MOHAMED BOUDIAF - M'SILA**

**FACULTY OF TECHNOLOGY**
**ELECTRONICS DEPARTMENT**
**N°: …./ 2025**

**DOMAIN: SCIENCE AND TECHNOLOGY**
**FILIERE: TELECOMMUNICATIONS**
**OPTION: ENGINEERING AND TELECOMMUNICATIONS**

**Dissertation Submitted in partial fulfilment of the requirements**
**For the Master professional Degree**
**By**:

Hachemi Manar Zahrat ELOla

Chennafi Karima

**Entitled**

# Sign Language Recognition System

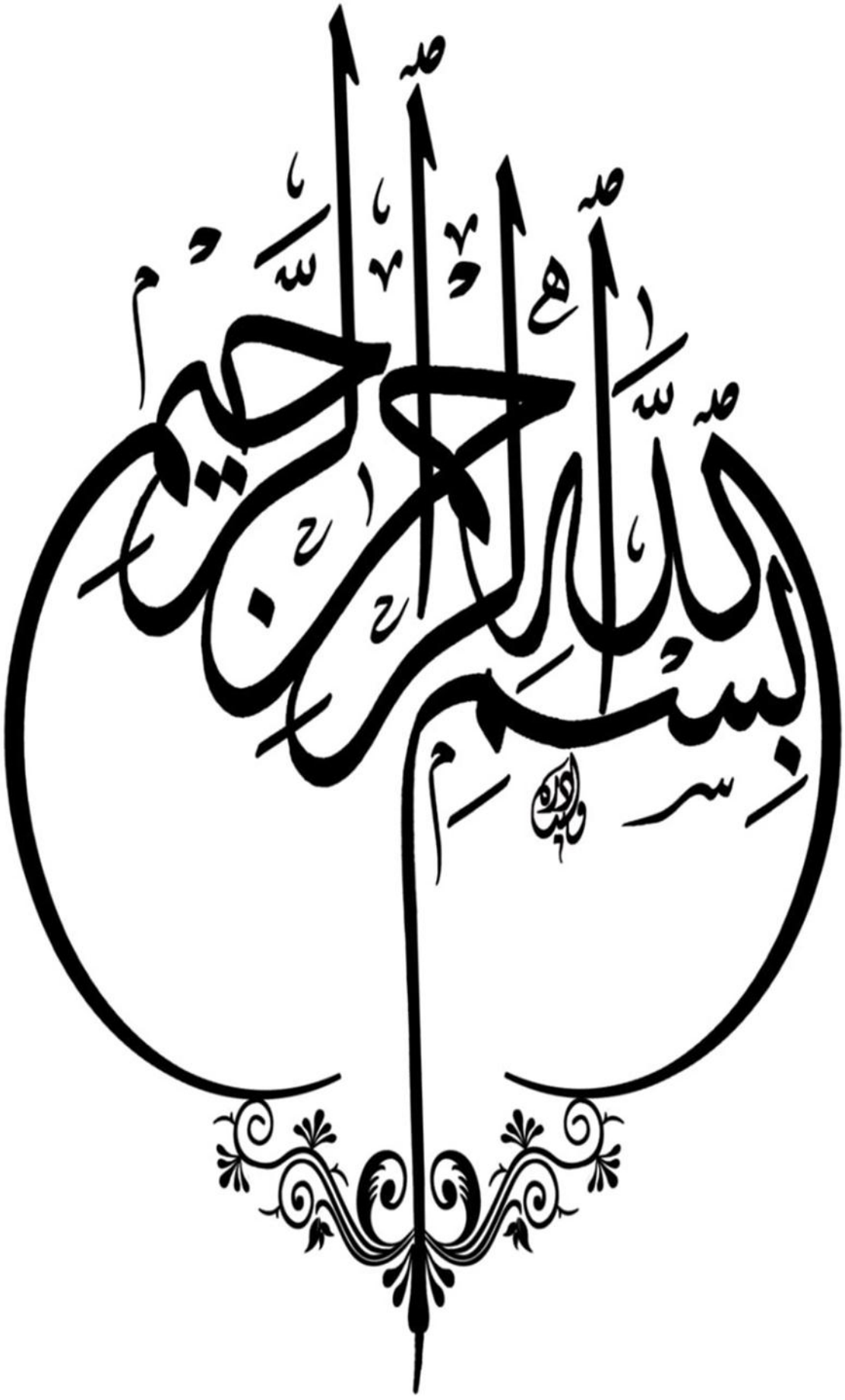
# Case Study : Algerian signs

**Master's degree under ministerial decree 1275**

**Presented on: 22th ,2025, in front of the jury composed of :**

| | | |
|---|---|---|
| Dr. BELOUTI Adel | President | Mohamed Boudiaf University - M'sila |
| Dr. BRIK Mourad | supervisor | Mohamed Boudiaf University - M'sila |
| Dr. ABED Ahcen | Examiner | Mohamed Boudiaf University - M'sila |
| DR. KHENNOUF Salah | CATI Representative | Mohamed Boudiaf University - M'sila |

**Academic year: 2024 / 2025**


بسم الله الرحمن الرحيم

# *Acknowledgments*

**يقول الله تعالى:**

**(رَبِّ أَوْزِعْنِي أَنْ أَشْكُرَ نِعْمَتَكَ الَّتِي أَنْعَمْتَ عَلَيَّ وَعَلَىٰ وَالِدَيَّ وَأَنْ أَعْمَلَ صَالِحًا تَرْضَاهُ وَأَدْخِلْنِي بِرَحْمَتِكَ فِي عِبَادِكَ الصَّالِحِينَ)**

**سورة النمل الآية 19.**

**ومصداقا لقول الرسول صلى الله عليه وسلم:" من لم يشكر الناس لم يشكر الله"**

*We would like to express our gratitude to the individuals who supported the completion of this thesis.*

*It is imperative to acknowledge the pivotal role of* ***Dr. Mourad Brik*** *in the genesis and development of this endeavor. Without his assistance and mentorship, the project would not have attained its full potential. We extend our profound gratitude to him for his exemplary supervision, patience, rigor, and availability during the preparation of this thesis.*

*Additionally, we would like to express our profound gratitude to the head of the electronics department at the University of M'sila,*

***Dr. Mostefa Tabbakh****, for his invaluable support.*

*We extend our profound gratitude to the esteemed professors and contributors who provided invaluable guidance through their insightful contributions, judicious counsel, and constructive critiques. Their willingness to engage with us in personal meetings and address our inquiries during the research process is particularly commendable.*

***Ola & karima***

# *Dedication*

*First of all, I thank Almighty God for giving me the courage, will, and patience to carry out this work despite all the difficulties I encountered.*

*To my beloved mother **Hadjira**, who endured countless sleepless nights and weary days, sacrificing her own comfort for mine since childhood, your unwavering love and support have shaped me into who I am today.*

*To my dear father **Abdelkrim**, for all the sacrifices you made and the hardships you endured, your unwavering support and guidance have been a constant source of inspiration for me.*

*To my siblings, **Rania**, **Nour**, **Raid ,**and **Sidra**, you have been the best siblings one could ever ask for, always there with unwavering support and love. And to the spirit of my dear my grandmother **Zerrouak Hada** and To my dear sister's husband **Hocine** for his constant support to me and my niece **Ania Celine**. And to my cat **Lucy**.*

*to my entire family, **Hachemi** and **Ogab**.*

*To all my brothers and Sisters not born of my mother, **Djihane**, **Zahra**, **Marwa**, **Assia**, **Chaimae**, **Imad**, **Zeyd**, **Riyadh**,**Ala**, **Khalil** ,**Taha**, for their support and their good and pleasant company in life.*

*Thanks for your unwavering support and delightful company, thank you for being there for me.*

*To the serene adventures, radiance surrounds, resonating in my life.*

*To my colleague in this project, **Karima** .*

***Ola***

# *Dedication*

*First of all, I thank Almighty God for giving me the courage, will, and patience to carry out this work despite all the difficulties I encountered. Indeed, the path of God is good.*

*To my dear mother,* ***Adiba*** *, for the sacrifices she made from my childhood until today,*

*To my dear father,* ***Aissa*** *, who has always supported and helped me face challenges, and who has always been and continues to be a role model for me.*

*To my siblings,* ***Bouchra****,* ***Ismaeil****,* ***Ibrahim****, and* ***Amine****, for being a refuge and a shelter from life's worries and problems.*

*to my entire family,* ***Chennafi*** *and* ***Haoues****.*

*To all my brothers and Sisters not born of my mother, for their support and their good and pleasant company in life. To The companions of the first step and the penultimate step. To those who in lean years were rain clouds.*

*Thanks for your unwavering support and delightful company, thank you for being there for me.*

*To the serene adventures, radiance surrounds, resonating in my life.*

*To my colleague in this project,* ***Ola.***

***Karima***

# *Table of contents*

Chapter 4. Implementation

## *List of Figures*

## *List of Tables*

## *List of Abbreviations*

| | |
|---|---|
| **ArSLR** | Arabic sign language recognition |
| **ASLR** | Automatic sign language recognition |
| **CNN** | Convolution Neural Network |
| **LSTM** | Long Short-Term Memory |
| **ArSL** | Arabic sign language |
| **ASL** | Algerian Sign Language |
| **AASL** | Alphabet of Algerian sign language |
| **ReLU** | Rectified Linear Unit |
| **GPT** | Generative Pre-Trained Transformers |
| **YOLO** | You Only Look Once |
| **VGG** | Visual Geometry Group |
| **AI** | Artificial Intelligence |
| **OCR** | Optical Charachter Recognition |
| **ML** | Machine Learning |
| **DL** | Deep Learning |
| **DNN** | Deep Neural Networks |
| **GPU** | Graphics Processing Unit |
| **LLMS** | Large Language Models |
| **RLHF** | Reinforcement Learning From Human Feedback |
| **RGB** | Red Green Blue |
| **FC** | Fully-Connected |
| **API** | Application Programming Interface |

# INTRODUCTION

# *INTRODUCTION*

Oral language is the most useful way for communication among individuals, encompassing a diverse range of dialects that differ from country to country and even from region to region. However, there is a segment of the population that struggles with spoken language due to issues related to pronunciation, hearing, or both, which leading to many challenges and obstacles that oral language cannot handle. Therefore, sign language emerges as a powerful communication tool and an efficient substitute for this state. Sign language operates as a communication system that utilizes hand and body movements, along with facial expressions, to convey meaningful messages instead of spoken words. This unique language provides a means to express their thoughts and engage with others. The sign language system varies across regions; therefore, we have dozens of sign languages, each specified by the local language, for example, there are many sign languages for the Arabic language, with each dialect characterized by unique signs. This diversity complicates the understanding process of the system, which is expressed through hand and body movements, coupled with facial expressions and other gestures that convey thoughts and emotions without relying on spoken words.

In this project, we aim to develop a system that can handle Arabic sign language by translating body movements including face and hands movements into comprehensible Arabic words. This task is challenging due to dialect variations and input quality. Therefore, we focused on using offline videos of individuals using sign language to convey messages, where our goal is to extract a complete and comprehensive Arabic phrase from each video, which may consist of more than one word.

Our work is organized into four chapters, where each one focuses on key aspects of our developed system.

- **The first chapter** presents the generalities of sign language recognition, which detailed the background of sign language, its recognition methods, and its application in the Arabic language.

- **The second chapter** discusses the Artificial intellegence and The Two types of Learnings (Machine Learning and Deep Learning ).

- **The third chapter** discusses the Algerian Sign Language ASL methods , specifically in the context of the Arabic language.

- **The fourth chapter** the Implementation.

# CHAPTER-1
# INTRODUCTION TO PATTERN RECOGNITION

## 1.1 Introduction

Sign language recognition represents the best solution and for the individuals with hearing impairments, and with technological advancements and the incorporation of artificial intelligence, researchers are actively investigating novel methods to improve the precision and effectiveness of sign language recognition systems. This chapter explores the basic concepts of sign language, detailing into the detection, identification, and categorization of sign language recognition methods, as well as the utilization of hand-sensing technology, with a specific focus on Arabic sign language recognition (ArSLR).

## 1.2 Sign language detection identification and recognition

Sign language, in general, is a visual language based on hand gestures used by humans to communicate. It involves coordinated movements of different parts of the body, including the hands, face, and body [1]. Sign language is recognized as the most structured form of gesture-based communication. Similar to spoken languages, sign language naturally develops within communities of individuals with hearing impairments, which evolve independently from the spoken language of the region. Each sign language has its own grammar and rules, all of which share the common feature of being visually perceived. Just as there are numerous spoken languages worldwide, there are also various sign languages used globally [2].

**Figure 1-1** Example of identification of the word ' الله ' in Arabic sign *gesture* [1].

Sign language detection involves determining whether a video from a diverse and unrestricted collection includes sign language content. This process has various applications, such as automatically tagging and categorizing videos, serving as an initial phase toward automatically captioning sign language videos, and facilitating the automatic initiation of Automatic Sign Language Recognition (ASLR) without relying on predetermined assumptions about the input

videos [3].

**Figure 1-2** Example of detection of the word ' الله ' in Arabic sign gesture[1].

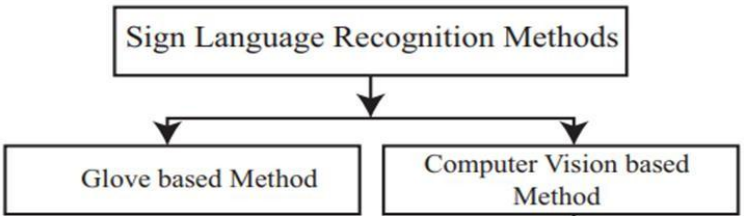
Sign language recognition is the final phase in the system which includes identifying and understanding sign language gestures, signs, and expressions using machine learning algorithms, deep learning models, or computer vision techniques to interpret them into letters or words which can be either in spoken language or text, which involves the classification and interpretation of hand gestures, finger positions, and body movements to accurately translate sign language into meaningful text or speech, enabling effective communication between individuals using sign language and those who do not understand sign language.[4]


**Figure 1-3** Example of recognition of the word ' الله ' in Arabic sign gesture[1].

## 1.3. Categorization of hand gesture recognition methods

Hand gesture recognition methods can be categorized into multiple methods based on different criteria. One of the main criteria is the type of input device used to capture the sign

language gestures. Depending on the input device, sign language recognition methods can be classified as glove-based, utilizing specialized gloves equipped with sensors for hand sensing, or vision-based (see **Figure 1-4**), using cameras or other visual sensors to capture hand movements and gestures. These categorizations are essential for understanding the diverse approaches in hand




sensing technology within sign language recognition.[5]

**Figure 1-4** Categorization of sign language recognition methods [2].

## 1.3.1 Computer vision-based method

The second major approach for detecting and recognizing hands involves utilizing standard video camera equipment, such as cameras or webcams, to capture a visual image of the user. Computer vision techniques are then employed to extract data about the hands from this image, including acquiring image series depicting hand postures or gestures. Subsequently, computer vision and image processing approaches are implemented to delineate the hand from the environment, elicit pertinent attributes, and discern the gestures. Benefits of this strategy encompass its naturalness, absence of intrusiveness, and affordability. Nevertheless, limitations emerge owing to susceptibility toward illumination variations, hindrances, and cluttered surroundings. Moreover, computer vision's capabilities are highlighted within this context. Vision-based methods can be further divided into two types as shown in **Figure 1-5**.

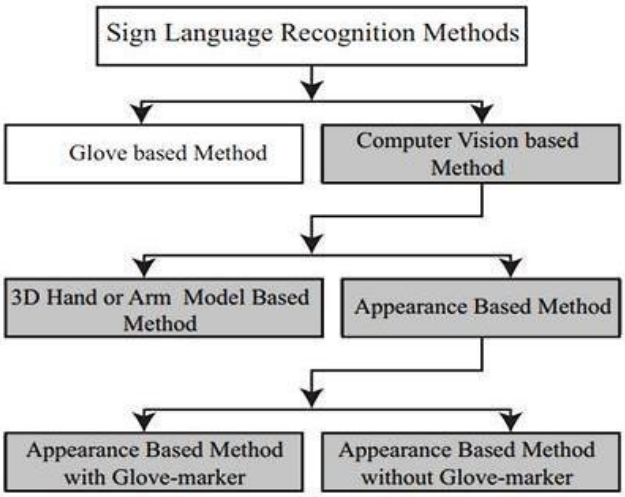


**Figure 1-5** Categorization of sign language recognition [2].

**A. 3D hand or arm model-based method:** Encompassing a subset of computer vision-driven strategies, this category utilizes a 3D representation of hands or arms to portray gestures. Model parameters are derived from pictures via tactics such as edge matching, form matching, or model adjustment. Strengths of this method comprise handling elaborate gestures and occlusions. Weaknesses incorporate elevated computational expenses and an extensive volume of instructional data required [6].

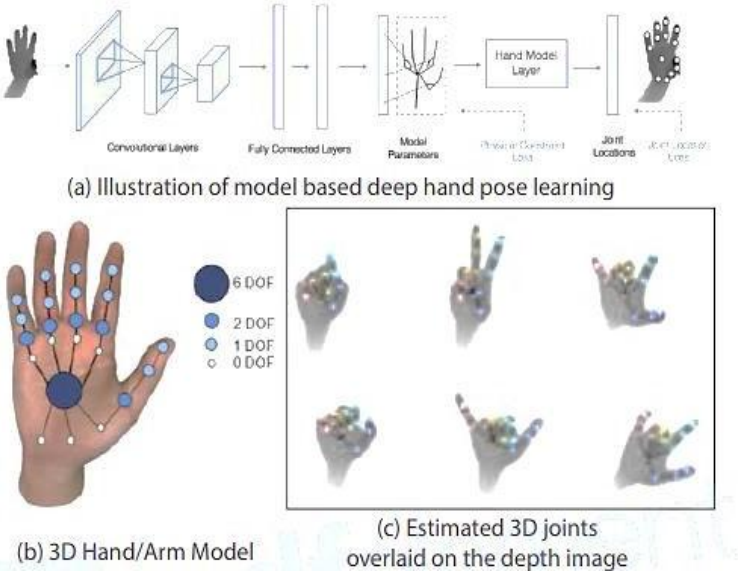



**Figure 1-6** An example of a 3D hand or arm model-based system[2].

**B. Appearance-based method:** The appearance-based method utilizes the appearance of the hand or the gesture as the feature and can be either with or without glove- markers. Glove-markers (**Figure** 1-8) are colored or illuminated markers that are attached to the glove to facilitate the segmentation **Figure** 1-7and feature extraction process. The proposed system in this paper uses an appearance-based method without glove-markers, which relies on skin-color detection and edge detection to segment the hand from the background and uses geometrical features to recognize the gestures [6]**.**

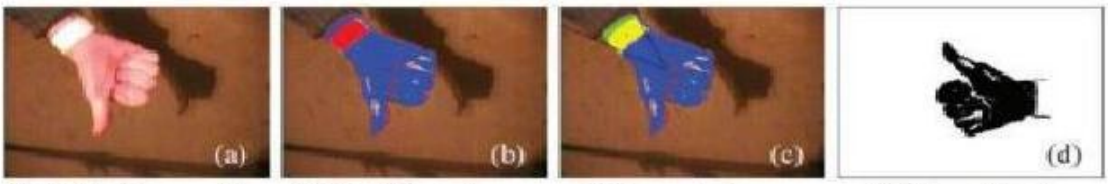


**Figure 1-7** An example of the hand-sign segmentation process using color wristband [2].


**Figure 1-8** An example uses of glove-markers [2].

## 1.4. Hand gesture recognition

Hand gesture recognition is a fundamental aspect of systems designed for sign language interpretation. This process entails the examination and comprehension of hand movements to decode and grasp the associated concepts within sign language. The efficacy of hand gesture recognition within these systems is underpinned by an amalgamation of tactile sensing devices, responsive control mechanisms, and algorithms dedicated to pattern identification. The findings from the aforementioned research underscore the proficiency and promise that hand sensing technologies hold in the domain of sign language recognition [7]. Hand gestures can be classified using the following approaches.

### A. Deep learning approaches

In the deep learning aspect, there are many approaches that are constantly used for hand gesture recognition. In the following, we present the most commonly used ones:

#### a. Convolutional Neural Network (CNN)

CNNs are deep learning models well-suited for image processing and are crucial in gesture recognition. They excel at extracting features from gesture images by utilizing multi- layer convolution and pooling operations to automatically learn spatial structures and local features. **Figure 1-9** depicts an architecture example of a CNN model used for hand gesture recognition.

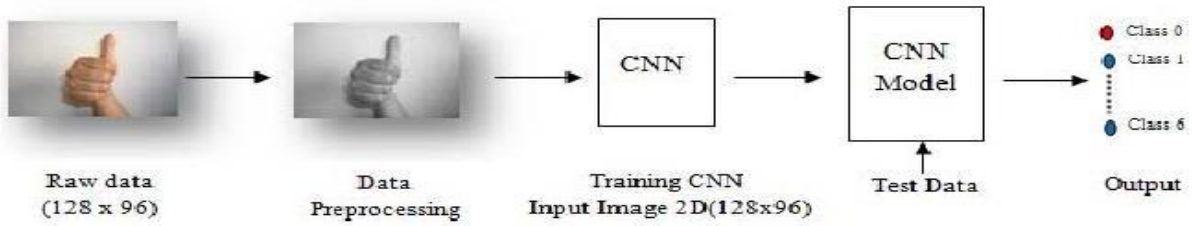



**Figure 1-9** The architecture of hand gesture recognition using a CNN model.[3]

Where CNN phase is described in the following elements:

- **Feature Extraction:** CNNs automatically extract features from images, such as edges, textures, and shapes, through convolution and pooling layers. These features effectively represent essential information in gestures, enabling accurate gesture recognition.
- **Hierarchical Representation:** Through successive convolution and pooling operations, CNNs construct a hierarchical representation of images. This representation captures features at various levels, enhancing the understanding of gesture structures and content for improved recognition accuracy.
- **Weight Sharing:** CNNs employ weight sharing in the convolutional layer, where the same set of weights is utilized across the entire image. This mechanism reduces model parameters, enhancing training efficiency and aiding in local feature extraction for gesture recognition.
- **Translation Invariance:** CNNs exhibit translation invariance, enabling them to recognize gestures even when translated within an image. This property enhances robustness in gesture recognition tasks, accommodating changes in gesture positions [8].

**a. Long Short-Term Memory (LSTM)**

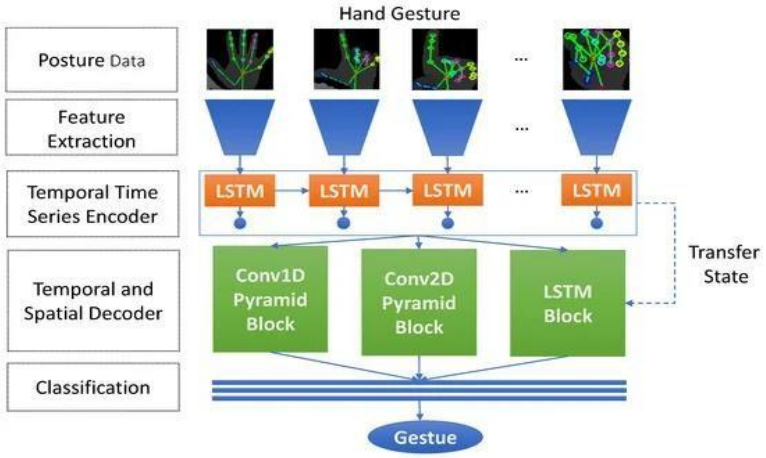



**Figure 1-10** The architecture of hand gesture recognition using a LSTM model.[4]

Where the key points of LSTM in gesture recognition include:

• **Processing Time Series Data:** LSTM effectively processes time series data, capturing temporal changes in gestures to understand timing features crucial for accurate recognition.

• **Long-Term Dependency Modeling:** LSTM mitigates issues like gradient vanishing or exploding gradients when dealing with long-term dependencies in gesture features. Its gating mechanism,

particularly the forgetting and input gates, enables robust modeling of long-term dependencies.

• **Modeling Context Information:** LSTM can model interrelated gestures by passing context information from previous gestures to subsequent ones. This capability enhances the understanding of gesture sequences, improving recognition accuracy[8].

## 1.5. Applications of hand gesture recognition

Hand gesture recognition technology has a wide array of applications, particularly in enhancing communication and learning.

## 1.6. Sign language recognition in Arabic language (ArSLR)

Arabic Sign Language (ArSL) is a comprehensive and organic means of communication adopted by the deaf population across Arab countries. The widespread unfamiliarity with ArSL exacerbates the marginalization of deaf people, further alienating them from societal participation. ArSL stands apart from spoken Arabic, showcasing unique grammatical.

### 1.6.1. Structural components of signs

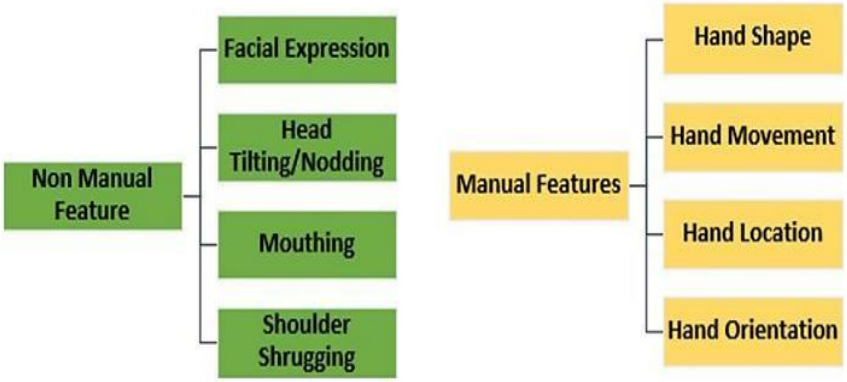



**Figure 1-11** Components of non-manual and manual features [5].

A. **Manual elements:** Those are based on the hand's shape, motion, position, and orientation. Some gestures are executed with one hand, while others require both hands. An illustration can showcase these manual elements through various gestures [9].

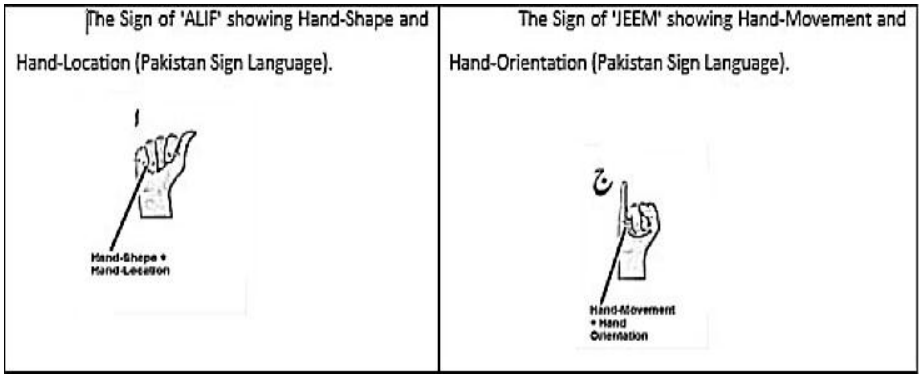



**Figure 1-12** Examples showing manual features [5].

**a. Gestures:** Gestures involving hand movements are known as dynamic signs, whereas those without hand movements are static signs. Gestures can also be categorized based on the number of hands used: single-handed for one hand, and double-handed for both hands. **Figure** 1-13 displays examples of single and double-handed gestures, both static and dynamic, with dynamic gestures illustrated through a sequence of frames [9].

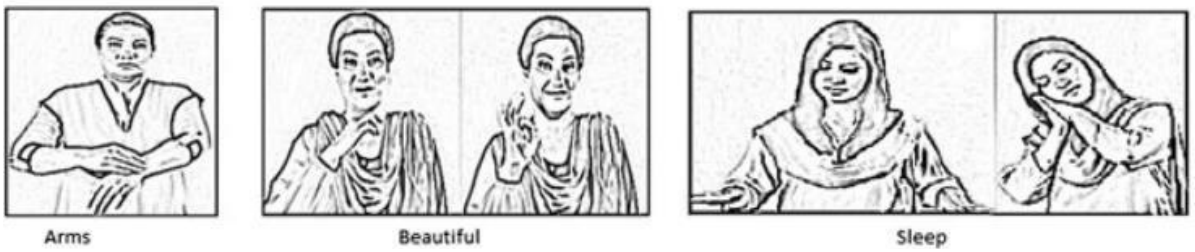



**Figure 1-13** An example of Single- and double handed gestures [5].

**A. Non-manual elements:** Encompass a range of facial expressions, head movements such as tilts and nods, shoulder movements, mouthing, and other actions that enhance the meaning of a sign. These non-manual markers typically accompany manual signs to convey complete information. Figure 1-14 can depict these non-manual elements, and another can demonstrate the differences between manual and non-manual gestures [9].

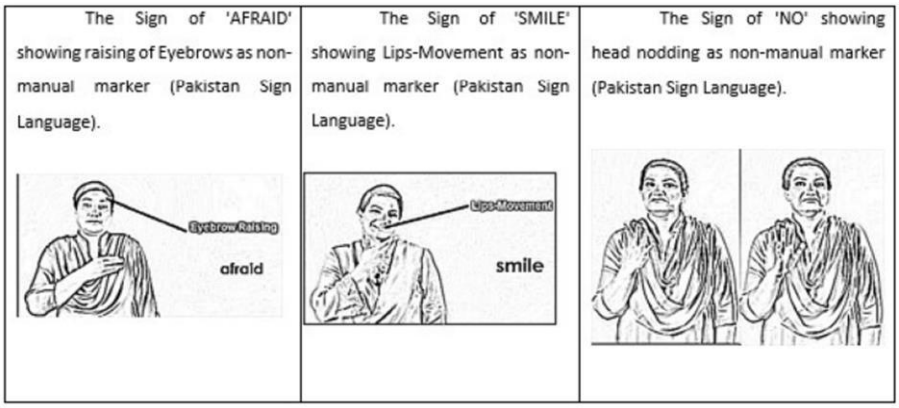



**Figure 1-14** An example showing non manual features [5].

## 1.6.2. Components of Arabic signs

In Arabic sign language, basic sign gestures consist of several key components including handshape, hand orientation, hand location, and movement. Understanding these components and measuring joint positioning and finger locations is essential for interpreting gestures and signs accurately in the context of Arabic sign language communication.

**A. Arabic hand shapes:** Arabic hand shapes are indeed a crucial component of ArSL. They comprise specific hand configurations and motions that express various words, letters, or ideas within the language. These hand shapes are integral to the grammar and lexicon of ArSL, allowing for effective communication within the deaf community in Arab regions [10].


**Figure 1-15** Hand Shapes Used for Arabic Alphabets.[6]

**B. Arabic hand orientations:** refers to the specific way that the hands are positioned while using ArSL. It plays a crucial role in conveying meaning and distinguishing between different signs, as the orientation of the hands can significantly alter the interpretation of a sign within the language.

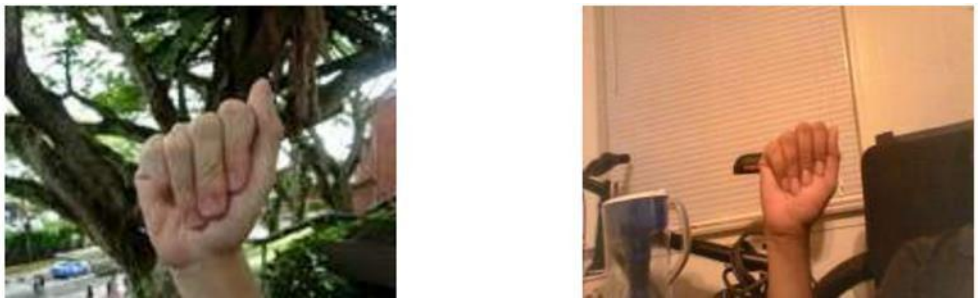
**C. Arabic hand locations:** Arabic hand locations refer to the specific areas of the body or space where the hands are placed during the formation of signs in ArSL. These locations are important in ArSL as they contribute to the clarity and understanding of the signs being conveyed. The placement can affect the meaning of a sign and is an integral part of the language's structure.

**D. Arabic hand motions:** Hand motions are indeed a vital component of ArSL. They refer to the various movements and gestures made by the hands while signing. These motions can include actions such as tapping, waving, pointing, or combining different movements to create meaning and convey complex concepts within the language. [10]

### 1.6.3. Sign language translation problems and challenges

### A. Uncontrolled Environment

Translation systems often struggle in environments with variable lighting, background noise, and other unpredictable factors that can affect the visibility and clarity of sign language gestures. Figure 1-16 presents samples of the challenges that can be encountered.



(a) Hand gesture with fair skin-tone and variable back-ground

(b) Hand gesture with fair skin-tone and poor lighting conditions

**Figure 1-16** Example Images of Hand Gestures with Variable Background and Lighting Conditions [7].

### A. Occlusion

Parts of the sign language gestures may be obscured by other body parts or objects, making it difficult for recognition systems to interpret the signs correctly. **Figure 1-17** and **Figure 1-18** present some cases where occlusion can lead to inaccurate detection.

**Figure 1-17** An Example of False Detection of Hand Region in Skin Based Detection Technique [7].

(A) ASL gesture for 'R' (B) ASL gesture for 'D'

**Figure 1-18** An example of Occlusion: R Gesture can Look like D in 2D Projection because of Occlusion [7].

## A. Low Inter-Class Variability

Signs that are similar to each other can be hard to distinguish, leading to errors in translation. This is especially challenging when signs differ only by slight variations in hand shape or movement. **Figure 1-19** presents an example of low inter-class variability, which can lead to misclassification of hand gestures.

(A) ASL Gesture for 'A'

(B) ASL Gesture for 'S'

**Figure 1-19** An example of Low Inter-Class Variability: A gesture can be Misclassified as S because of Low Inter-Class Variability [7].

## 1.7. The difference between Arabic sign language and Algerian sign language

Arabic Sign Language is different from Algerian Sign Language. Although there are some similarities due to shared cultural and religious factors, each Arab country has its own sign language influenced by local dialects, culture, and history.

Arabic Sign Language is a general term for the sign languages used in the Arab world, but in reality, there is great diversity among the sign languages of Arab countries. For example, Algerian Sign Language (ALSL) differs from Egyptian, Saudi, or Tunisian Sign Language.

Algerian Sign Language developed independently and is influenced by the Arabic spoken in Algeria (the Algerian dialect) and some French influences due to colonial history. Therefore, the signs used in Algeria may differ from those used in other Arab countries.

On the other hand, it is a sign language specific to Algeria and was officially recognized in 2002. ALSL belongs to the French Sign Language family, making it distinct from other Arabic sign languages.

## 1.8. the main differences between Arabic and Algerian Sign Language

Arabic Sign Language differs from Algerian Sign Language in several aspects:

**1. Cultural basis:** Arabic Sign Language is similar in some Arab countries due to a shared cultural heritage, while Algerian Sign Language has been influenced by local French and Algerian culture.

**2. Linguistic influence:** Algerian Sign Language shows influence from French, making it distinct from other Arabic Sign Languages.

**3. Local usage:** Each sign language is primarily used in its own country and relies on local signs

and expressions.

## 1.9. Conclusion

In conclusion, the recognition of sign language is an important field that has to be developed and expanded more for the benefit of the individuals with disabilities, which consists of viable strategies for enhancing hearing-impaired or deaf people's communication and increasing inclusivity and accessibility for sign language users. The upcoming chapter will examine current studies aimed at Artificial intelligence in the multimedia.

# CHAPTER-2
# ARTIFICIAL INTELLIGENCE IN THE MULTIMEDIA

## 2.1 Introduction

Artificial intelligence (AI) is a set of technologies that enable computers to perform a variety of advanced functions, including the ability to visualize, understand and translate spoken and written language, analyze data, make recommendations and more.

AI is at the heart of innovation in modern computing. It enables users and companies to unlock value. For example, optical character recognition (OCR) uses AI to extract text and data from images and documents, transforming unstructured content into structured data suitable for business, and unlocking valuable insights.

## 2.2 Artificial Intelegence

Artificial intelligence (AI) refers to the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making. It is a field of research in computer science that develops and studies methods and software that enable machines to perceive their environment and use learning and intelligence to take actions that maximize their chances of achieving defined goals. Such machines may be called AIs.

High-profile applications of AI include advanced web search engines (e.g., Google Search); recommendation systems (used by YouTube, Amazon, and Netflix); virtual assistants (e.g., Google Assistant, Siri, and Alexa); autonomous vehicles (e.g., Waymo); generative and creative tools (e.g., ChatGPT and AI art); and superhuman play and analysis in strategy games (e.g., chess and Go). However, many AI applications are not perceived as AI: "A lot of cutting edge AI has filtered into general applications, often without being called AI because once something becomes useful enough and common enough it's not labeled AI anymore."

Various subfields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include learning, reasoning, knowledge representation, planning, natural language processing, perception, and support for robotics. General intelligence the ability to complete any task performed by a human on an at least equal level is among the field's long-term goals. To reach these goals, AI researchers have adapted and integrated a wide range of techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, operations research, and economics. AI also draws upon psychology, linguistics, philosophy, neuroscience, and other fields.

Artificial intelligence was founded as an academic discipline in 1956, and the field went through multiple cycles of optimism throughout its history, followed by periods of disappointment and loss of funding, known as AI winters. Funding and interest vastly increased after 2012 when deep learning outperformed previous AI techniques. This growth accelerated further after 2017 with the transformer architecture, and by the early 2020s many billions of dollars were being invested in AI and the field experienced rapid ongoing progress in what has become known as the AI boom. The emergence of advanced generative AI in the midst of the AI boom and its ability to create and modify content exposed several unintended consequences and harms in the present and raised concerns about the risks of AI and its long-term effects in the future, prompting discussions about regulatory policies to ensure the safety and benefits of the technology. [11]

## 2.3 General intelligence

A machine with artificial general intelligence should be able to solve a wide variety of problems with breadth and versatility similar to human intelligence.[12]

## 2.4 Learning

Machine learning is the study of programs that can improve their performance on a given task automatically. It has been a part of AI from the beginning.

There are several kinds of machine learning. Unsupervised learning analyzes a stream of data and finds patterns and makes predictions without any other guidance. Supervised learning requires labeling the training data with the expected answers, and comes in two main varieties: classification (where the program must learn to predict what category the input belongs in) and regression (where the program must deduce a numeric function based on numeric input).

In reinforcement learning, the agent is rewarded for good responses and punished for bad ones. The agent learns to choose responses that are classified as "good".. Transfer learning is when the knowledge gained from one problem is applied to a new problem. Deep learning is a type of machine learning that runs inputs through biologically inspired artificial neural networks for all of these types of learning.

Computational learning theory can assess learners by computational complexity, by sample complexity (how much data is required), or by other notions of optimization. [13]

### 2.4.1 Machine Learning (ML)

Machine learning (ML) is a discipline of artificial intelligence (AI) that provides machines with the ability to automatically learn from data and past experiences while identifying patterns to

make predictions with minimal human intervention.

Machine learning methods enable computers to operate autonomously without explicit programming. ML applications are fed with new data, and they can independently learn, grow, develop, and adapt.

Machine learning derives insightful information from large volumes of data by leveraging algorithms to identify patterns and learn in an iterative process. ML algorithms use computation methods to learn directly from data instead of relying on any predetermined equation that may serve as a model.

The performance of ML algorithms adaptively improves with an increase in the number of available samples during the 'learning' processes. For example, deep learning is a sub-domain of machine learning that trains computers to imitate natural human traits like learning from examples. It offers better performance parameters than conventional ML algorithms.

While machine learning is not a new concept – dating back to World War II when the Enigma Machine was used – the ability to apply complex mathematical calculations automatically to growing volumes and varieties of available data is a relatively recent development.

Today, with the rise of big data, IoT, and ubiquitous computing, machine learning has become essential for solving problems across numerous areas, such as :

- Computational finance (credit scoring, algorithmic trading)
- Computer vision (facial recognition, motion tracking, object detection)
- Computational biology (DNA sequencing, brain tumor detection, drug discovery)
- Automotive, aerospace, and manufacturing (predictive maintenance)
- Natural language processing (voice recognition) [14]

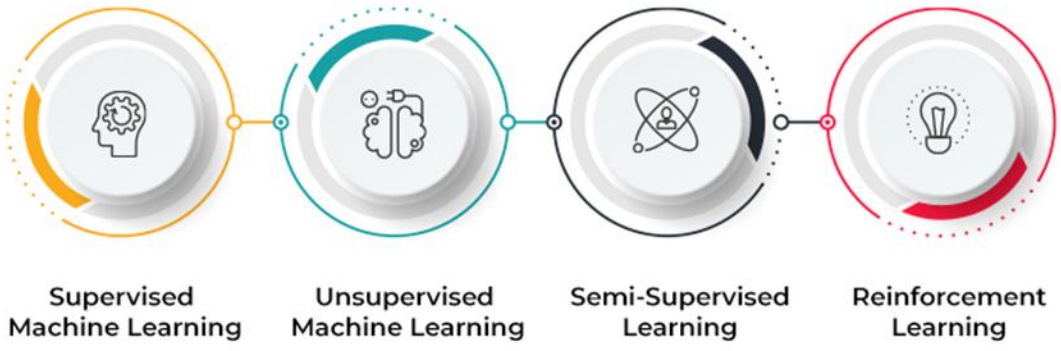



**Figure 2-1** Types of Machine Learning [8]

### 2.4.1.1 supervised learning

Supervised machine learning requires labelled input and output data during the training phase of the machine learning model lifecycle. This training data is often labelled by a data scientist in the preparation phase, before being used to train and test the model. Once the model has learned the relationship between the input and output data, it can be used to classify new and unseen datasets and predict outcomes.

The reason it is called supervised machine learning is because at least part of this approach requires human oversight. The vast majority of available data is unlabelled, raw data. Human interaction is generally required to accurately label data ready for supervised learning. Naturally, this can be a resource intensive process, as large arrays of accurately labelled training data is needed.

Supervised machine learning is used to classify unseen data into established categories and forecast trends and future change as a predictive model. A model developed through supervised machine learning will learn to recognise objects and the features that classify them. Predictive models are also often trained with supervised machine learning techniques. By learning patterns between input and output data, supervised machine learning models can predict outcomes from new and unseen data. This could be in forecasting changes in house prices or customer purchase trends.

Supervised machine learning is often used for:

- Classifying different file types such as images, documents, or written words.
- Forecasting future trends and outcomes through learning patterns in training data.

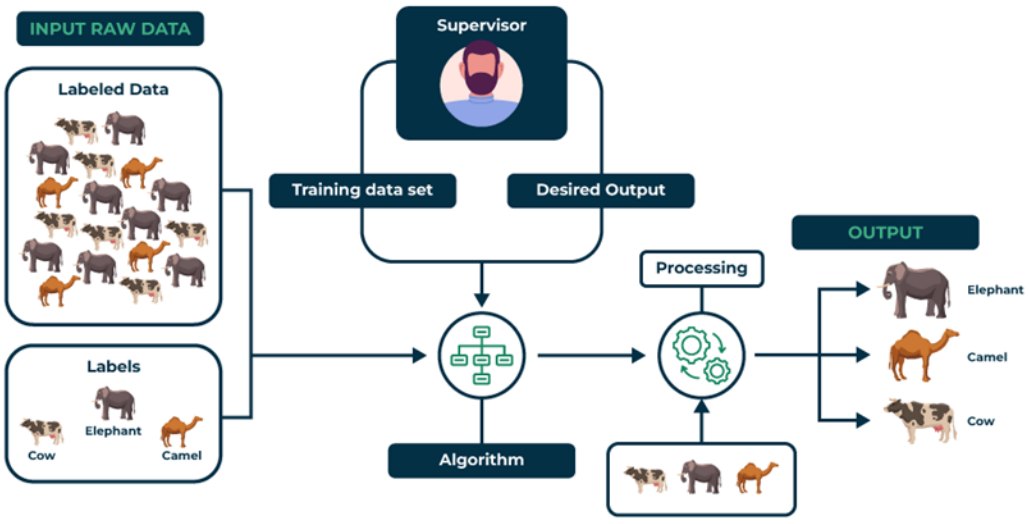



**Figure 2-2:** Supervised Learning [9]

### 2.4.1.2 unsupervised learning

Unsupervised machine learning is the training of models on raw and unlabelled training data. It is often used to identify patterns and trends in raw datasets, or to cluster similar data into a specific number of groups. It's also often an approach used in the early exploratory phase to better understand the datasets.

As the name suggests, unsupervised machine learning is more of a hands-off approach compared to supervised machine learning. A human will set model hyperparameters such as the number of cluster points, but the model will process huge arrays of data effectively and without human oversight. Unsupervised machine learning is therefore suited to answer questions about unseen trends and relationships within data itself. But because of less human oversight, extra consideration should be made for the explainability of unsupervised machine learning.

The vast majority of available data is unlabelled, raw data. By grouping data along similar features or analyzing datasets for underlying patterns, unsupervised learning is a powerful tool used to gain insight from this data. In contrast, supervised machine learning can be resource intensive because of the need for labelled data.

Unsupervised machine learning is mainly used to:

- Cluster datasets on similarities between features or segment data
- Understand relationship between different data point such as automated music recommendations
- Perform initial data analysis

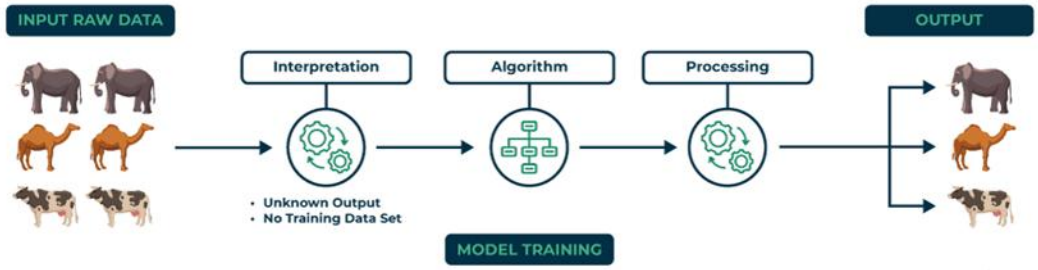



**Figure2-3:** Unsupervised Learning [10]

## 2.4.2 Deep Learning (DL)

Deep learning uses several layers of neurons between the network's inputs and outputs. The multiple layers can progressively extract higher-level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits, letters, or faces.

Deep learning has profoundly improved the performance of programs in many important

subfields of artificial intelligence, including computer vision, speech recognition, natural language processing, image classification, and others. The reason that deep learning performs so well in so many applications is not known as of 2021. The sudden success of deep learning in 2012–2015 did not occur because of some new discovery or theoretical breakthrough (deep neural networks and backpropagation had been described by many people, as far back as the 1950s) but because of two factors: the incredible increase in computer power (including the hundred-fold increase in speed by switching to GPUs) and the availability of vast amounts of training data, especially the giant curated datasets used for benchmark testing, such as ImageNet.[15]

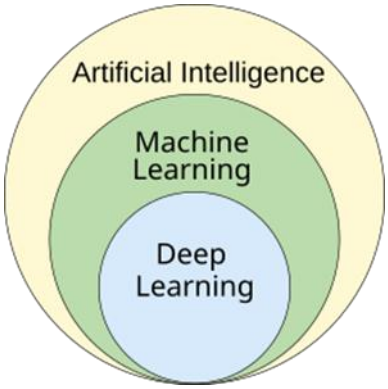



**Figure2-4:** Types of Machine Learning [11]

## 2.5 GPT

Generative pre-trained transformers (GPT) are large language models (LLMs) that generate text based on the semantic relationships between words in sentences. Text-based GPT models are pretrained on a large corpus of text that can be from the Internet. The pretraining consists of predicting the next token (a token being usually a word, subword, or punctuation). Throughout this pretraining, GPT models accumulate knowledge about the world and can then generate human-like text by repeatedly predicting the next token. Typically, a subsequent training phase makes the model more truthful, useful, and harmless, usually with a technique called reinforcement learning from human feedback (RLHF). Current GPT models are prone to generating falsehoods called "hallucinations". These can be reduced with RLHF and quality data, but the problem has been getting worse for reasoning systems. Such systems are used in chatbots, which allow people to ask a question or request a task in simple text.

Current models and services include Gemini (formerly Bard), ChatGPT, Grok, Claude, Copilot, and LLaMA. Multimodal GPT models can process different types of data (modalities) such as images, videos, sound, and text. [16]

## 2.6 Yolo ( You Only Look One )

### 2.6.1 Definition

YOLO (You Only Look Once) models are real-time object detection systems that identify and classify objects in a single pass of the image. In other words, the model only looks at the image once and from this 'single pass' is able to identify objects in the image. This is different from previous models that required multiple passes to process an image. Since this process happens in real-time, it works at incredible speed.

Object detection means that YOLO can not only pinpoint where an object is in an image but also what it is. This works by passing the image through a neural network that allows the model to detect all objects simultaneously. Using grid-based prediction and bounding box prediction, the model is able to understand if objects fall within a certain cell or bounding box. Additionally, YOLO uses class probability to predict how to classify objects, which determines what the object actually is.[17]

### 2.6.2 The evolution of the Yolo table

| YOLO Version | Year | Key Advancements |
|---|---|---|
| YOLOv1 | 2015 | Introduction of real-time object detection using a grid-based approach |
| YOLOv2 | 2016 | Integration of anchor boxes, pyramid feature networks, and multi-scale prediction |
| YOLOv3 | 2018 | Improvements in accuracy and speed with the introduction of Darknet-53 and multiple detection scales |
| YOLOv8 | 2021 | Cutting-edge advancements in real-time object detection with improved accuracy and speed |

**Table 2-1 :** The evolution of the Yolo [1]

### 2.6.3 The purpose of Yolo

YOLO (You Only Look Once) is a real-time object detection algorithm developed by Joseph Redmon and Ali Farhadi in 2015. It is a single-step object detector that uses a convolutional neural network (CNN) to predict bounding boxes and class probabilities of objects in input images . [18]

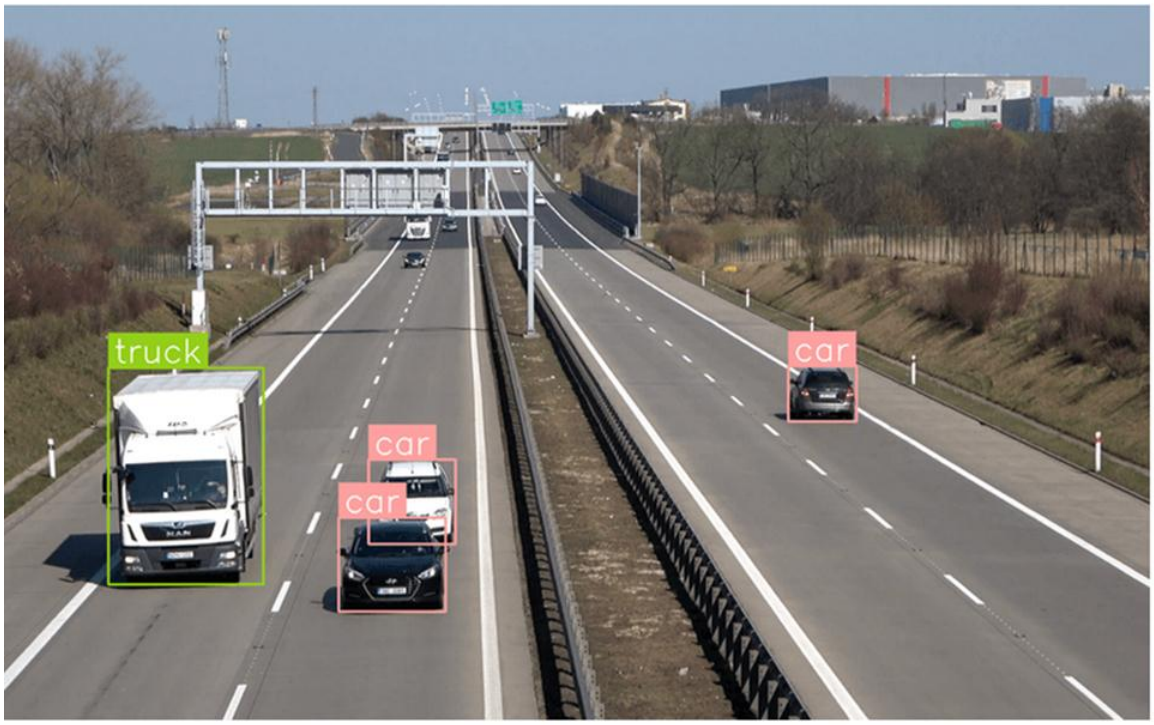



**Figure2-5 :** yolov8 detection

## 2.6.4 The workflow of YOLO in the objective detection

The most significant feature of YOLO class models is that they simplify the two steps in object detection into one step. The conventional approach for object detection is to first identify candidate regions of suitable size, and then classify the candidate regions . YOLO directly scans the entire image and provides both category and location information in one step . YOLO first divides the image into several small pieces, and then creates bounding boxes for each small piece. The bounding boxes can be large or small, and may even exceed the size of the original small block, but YOLO only requires the object centre of the bounding box to be in the small block. In this way, YOLO transforms the problem into a regression problem. Fig 2-6 shows the workflow of YOLO in the objective detection.[19]

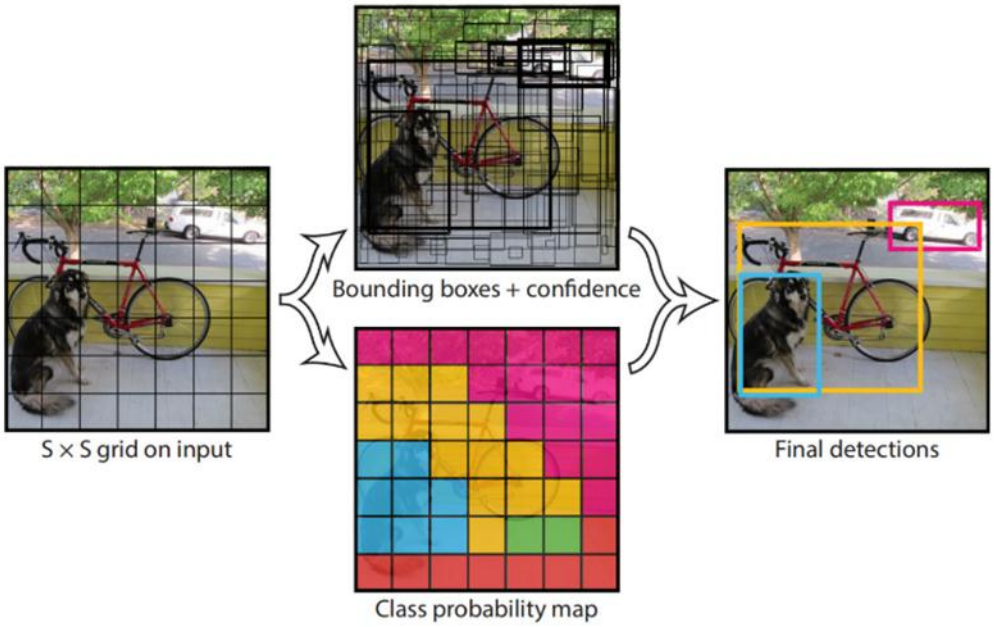



**Figure2-6:** The workflow of YOLO in the objective detection [13]

## 2.7 Vgg (Visual Geometry Group)

### 2.7.1 Vgg Definition

Visual Geometry Group (VGG) A significant feature of the VGG model is that it uses smaller convolutional kernels on top of the previously proposed models. There are many benefits in this way, one of which is making the model deeper. The process of training convolutional neural networks is the process of continuously extracting abstract features through convolutional kernels, in other words, extracting useful information from images. A deeper network structure means that the network has the ability to extract more important features . For example, VGG uses three 3x3 convolutional kernels instead of the 7x7 convolutional kernels in AlexNet, which changes the model from only extracting features once to three times. Moreover, these three times of extraction are continuously refined, meaning that the latter abstraction continues on top of the previous one, resulting in higher-level abstractions. Another advantage is that it has fewer parameters while ensuring receptive fields, which means it is easier to train . For example, a 7x7 convolutional kernel has 49 parameters, while three 3x3 convolutional kernels only have 09 parameters. At the same time, compared to large convolutional kernels, small convolutional kernels need to be moved more times for images of the same size in both step sizes of 1, which means they can better read the details of images [20] . The following Fig 2 shows the structure of the classic VGG16 model.

## 2.7.2 VGG16 purpose

VGG16 is object detection and classification algorithm which is able to classify 1000 images of 1000 different categories with 92.7% accuracy. It is one of the popular algorithms for image classification and is easy to use with transfer learning. [21].

## 2.7.3 VGG16 Architecture

The VGG-16 architecture is a deep convolutional neural network (CNN) designed for image classification tasks. It was introduced by the Visual Geometry Group at the University of Oxford. VGG-16 is characterized by its simplicity and uniform architecture, making it easy to understand and implement.

The VGG-16 configuration typically consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. These layers are organized into blocks, with each block containing multiple convolutional layers followed by a max-pooling layer for downsampling.[22]

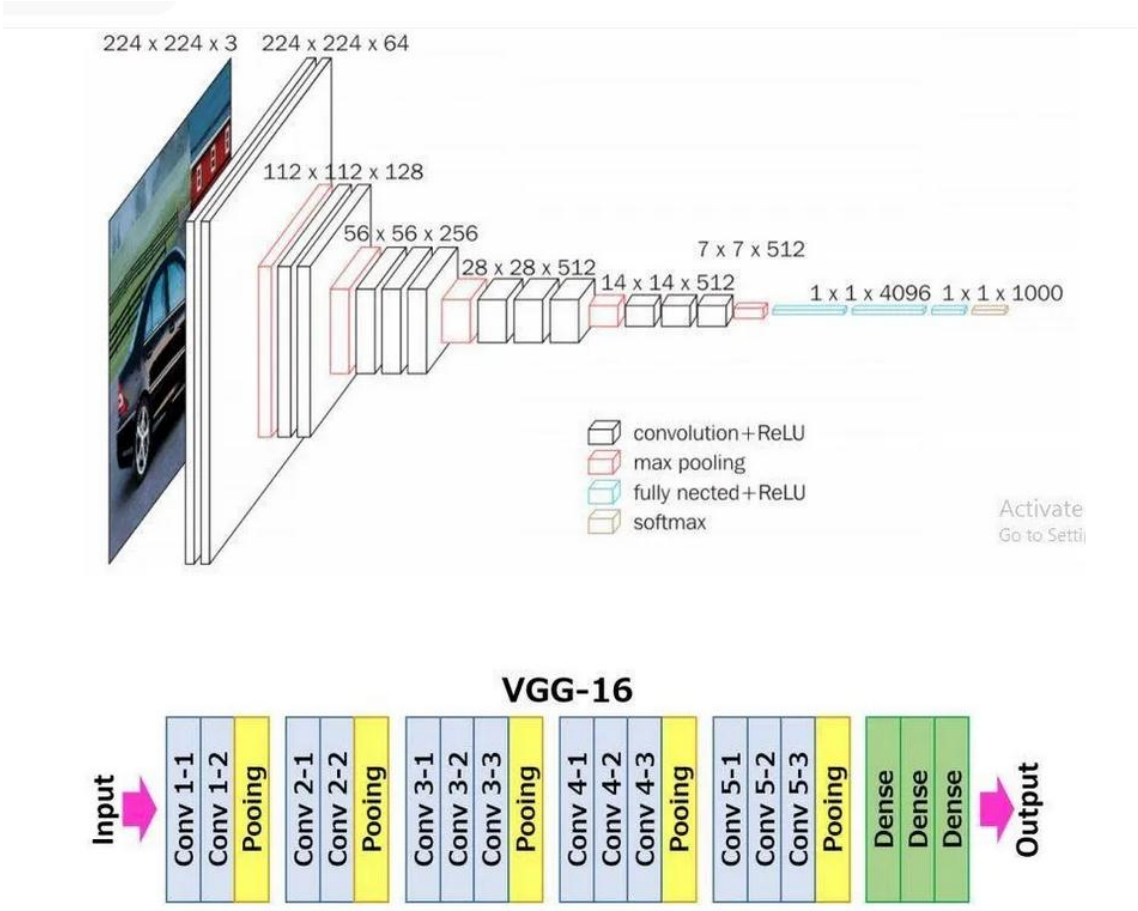



**Figure2-7:** VGG16 Architecture [14]

• The 16 in VGG16 refers to 16 layers that have weights. In VGG16 there are thirteen convolutional layers, five Max Pooling layers, and three Dense layers which sum up to 21 layers but it has only sixteen weight layers i.e., learnable parameters layer.

• VGG16 takes input tensor size as 224, 244 with 3 RGB channel

• Most unique thing about VGG16 is that instead of having a large number of hyper-parameters they focused on having convolution layers of 3x3 filter with stride 1 and always used the same padding and maxpool layer of 2x2 filter of stride 2.

• The convolution and max pool layers are consistently arranged throughout the whole architecture

• Conv-1 Layer has 64 number of filters, Conv-2 has 128 filters, Conv-3 has 256 filters, Conv 4 and Conv 5 has 512 filters.

• Three Fully-Connected (FC) layers follow a stack of convolutional layers: the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. [23]

## 2.8 Conclusion

Artificial Intelligence complements human intelligence to produce goods and services, and thus wealth. From this perspective, the role of the human being in the quest for efficiency becomes central and strategic. We will present in the next chapter Algerian sign language ASL.

# CHAPTER-3
# ALGERIAN SIGN LANGUAGE

## 3.1 Introduction

The field of sign language recognition is a well-established area of research with a diverse range of implementation methods. In this chapter, we endeavored to find the largest possible number of studies in Arabic language, considering the significant challenges associated with studying this field, especially when focusing solely on the Arabic language. Hereafter, we have gathered the most recent and important works focused on Arabic sign language, irrespective of its dialects.

## 3.2 Basic methodology

All the works that are presented here are focused on Arabic language generation from variation type of signs which can be extracted from either video or image data. The figure below represents the basic steps of any approach developed under the vision-based method or with sensors (for more details, see Chapitre 1).

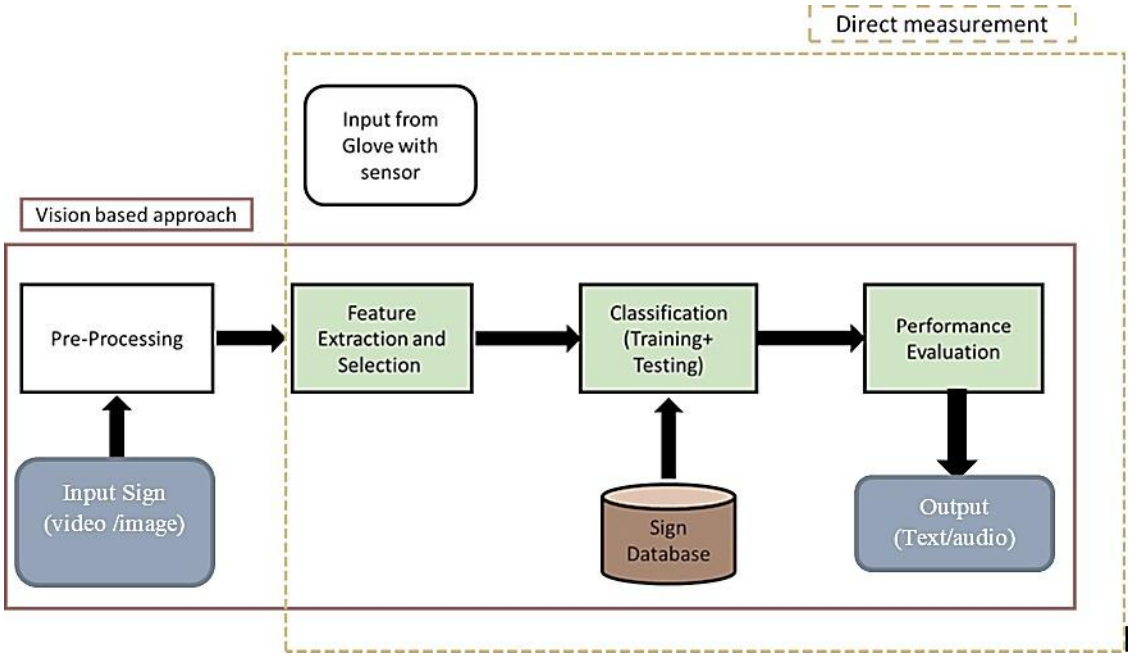



**Figure 3-1** System's general block design for recognizing sign language.[ 15]

The presented works have been divided into three main groups: image processing/statistical modeling-based recognition, classic machine learning-based

recognition, and deep learning-based recognition.

## 3.3 Algerian Sign Language

### 3.3.1 Definition

Algerian Sign Language (ASL) is the sign language used by deaf people and their relatives in Algeria to understand each other through signs. Algerian Sign Language is officially recognized by the law of May 8, 2002 as the first language of the deaf-mute community in Algeria, which is the only country in the Arab world and Africa to officially recognize sign language [11 !!!].

ASL is entirely biased on gestures (signs), each sign is made by means of different parts of the body, the hand(s), the face, the shoulder, ... or by the whole body. Intuitively speaking, ASL is a language like all other languages. Indeed, it has a vocabulary, and an organized syntax just like spoken languages. So, to learn, and understand ASL, then quite simply one must know its alphabet. In fact, every sign of the LSA alphabet is galore by one or two particular posture(s) of the hand [24].

### 3.3.2 Types of Signs

A sign is a class of gestures that depends or not on a certain duration in time. We can first classify gestures according to the parts of the body involved. We generally distinguish three types of gestures [25]:

#### 3.3.2.1 Hand and Arm Gestures

They form the main category of interactive gestures. The hand allows precise and complex gestures to be performed. Research around these gestures mainly concerns the recognition of hand positions, the interpretation of sign language and allowing the manipulation and interaction with data or elements of an environment.

#### 3.3.2.2 Head and Face Gestures

Few head gestures have specific meaning; head orientation is very useful for field of view detection.

### 3.3.3 Alphabet of Algerian Sign Language (AASL)

In Algerian Sign Language, there are 42 signs of the Algerian alphabet, among which there are 37 static signs and 5 dynamic signs. In fact, these signs are represented by a single hand (see Figure 3-2). Furthermore, each of the static signs is determined by means of two parameters which are: configuration and orientation.

**Figure 3-2:** Alphabet of Algerian Sign Language (ASL) [16].

## 3.4 Characteristics of Arabic dataset used

For developing our system using deep learning techniques, we attempted to find the most suitable dataset that meets our needs. However, there was a significant absence of datasets where the input is video and in Arabic language. We found only two datasets where video input was available, and currently, only one of them is accessible. This limited our options to develop our models based on this dataset.

### 3.4.1 Our used dataset

Basing on the provided dataset (ArabSign-A dataset), we created a new alternative for our system, where we extracted the segments of each video, with each segment representing one word. We manually divided 30 videos of a sentence into segments, resulting in 5 distinct words, each represented by 30 videos. Therefore, rather than working on a full sentence or full video, we focused on individual words in Arabic sign language. Figure 3-3 presents the new generation of our dataset used.

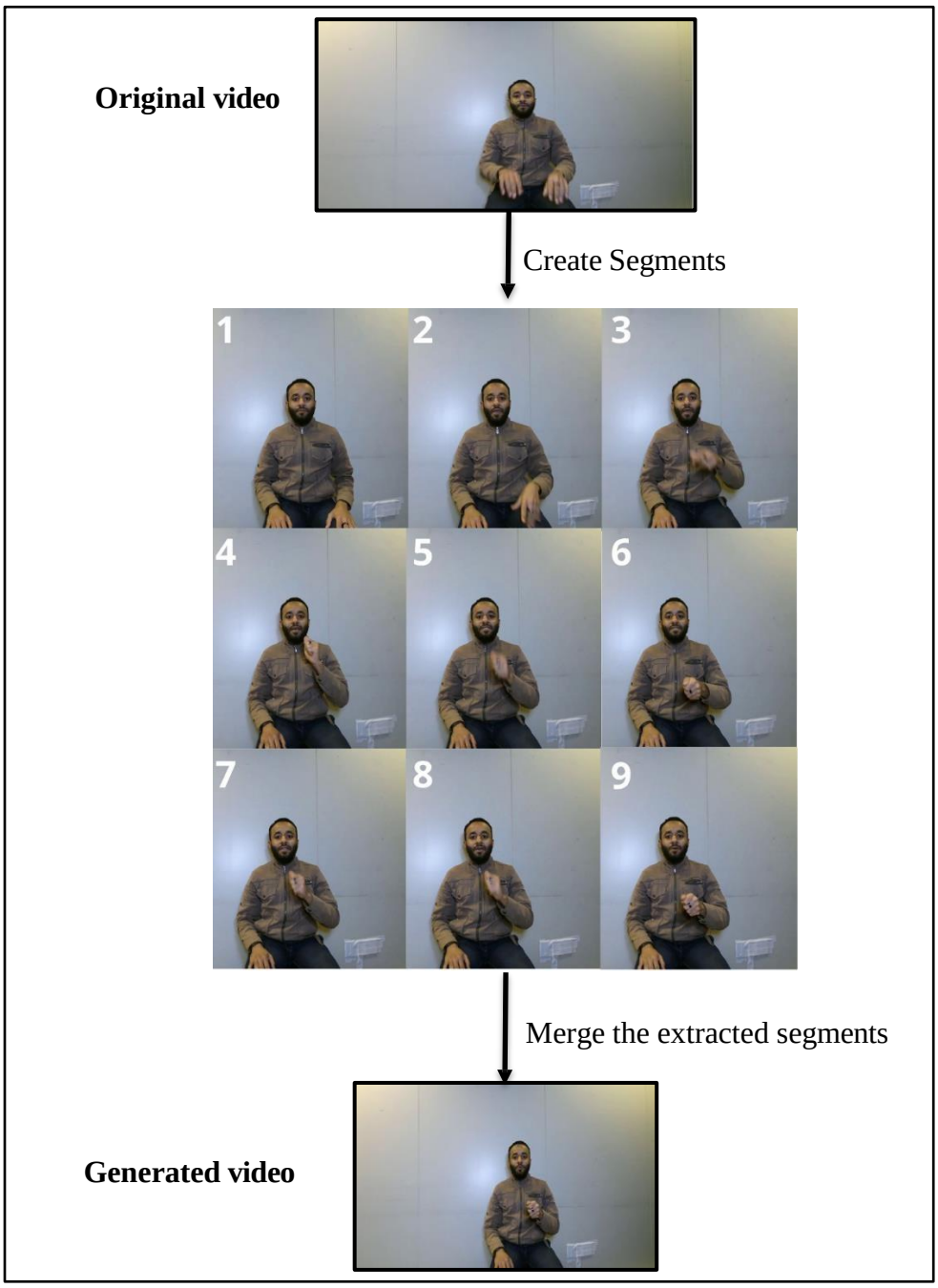



**Figure 3-3** A sample of the new generation dataset based on ArabSign-A dataset.

## 3.5 Proposed system architecture

The proposed approach of our system is presents in Figure 3-4 which illustrates the basic phases of the sing language recognition for translating the provided sign language into Arabic words. The proposed system consists of the following phases:

• Preprocessing phase: In this stage, we prepare the video, extract a set of frames from it, normalize it, and optimize its lighting after resizing it

• Object Localization and Extraction: This stage complements the previous stage by locating the primary objects for badge recognition, including hands and face.

• Holistic detection: After locating the objects in this stage, we detail the extracted objects for more information about the face, hands, and body features.

• Segmentation phase: At this stage, the video is segmented into 1-second clips in order to deal with each tag separately

• Recognition phase: In this phase, we use deep learning techniques to recognize Arabic sign language using four separate models that we have developed: CNN, LSTM, YOLOv8, and a hybrid CNN-LSTM model.

This systematic approach ensures the accurate identification and translation of sign language into

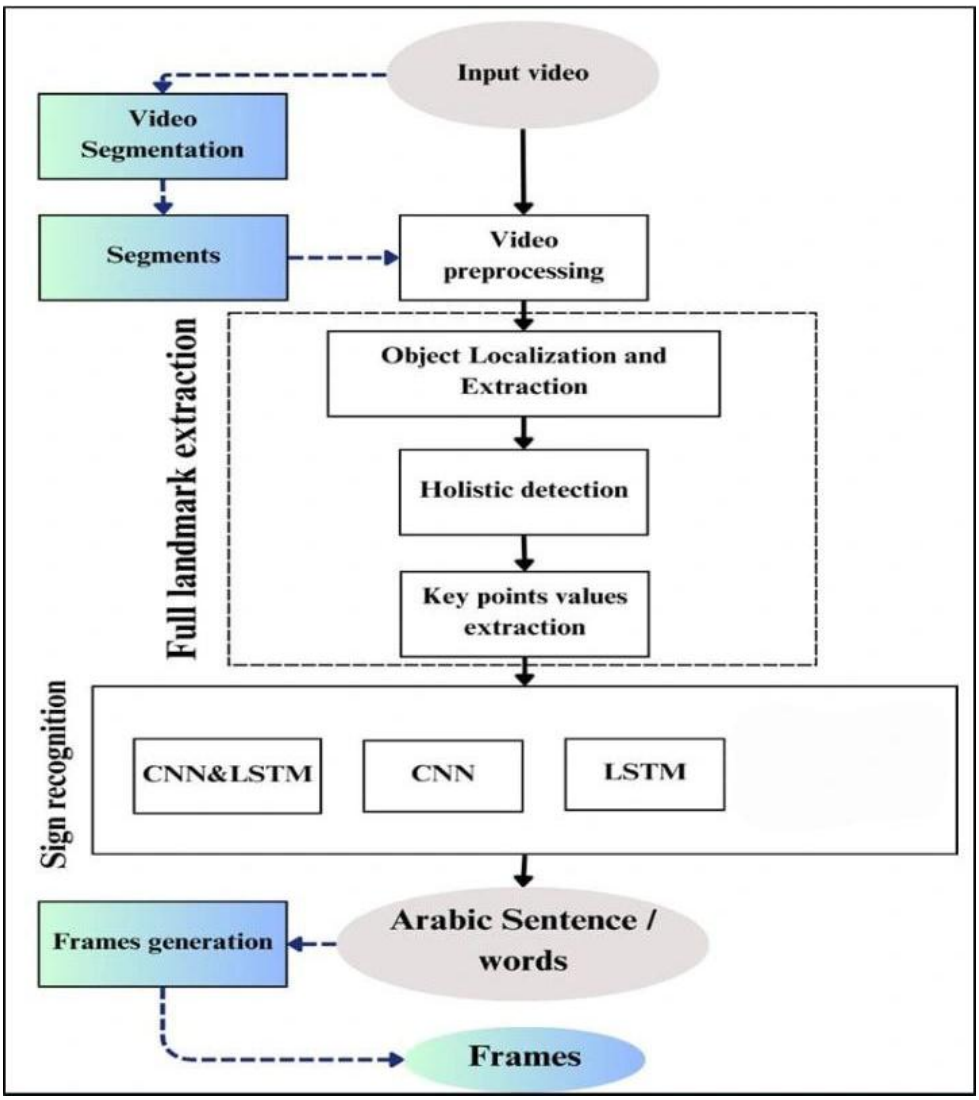



**Figure 3-4** The architecture of the proposed Arabic sign language recognition system

### 3.5.1 Video preprocessing

The first phase in our system is preprocessing the provided video to ensure accurate results in subsequent phases of our proposed system. The following pseudocode represents a generalized preprocessing pipeline of this phase:

```
function preprocessVideo:
  Input= video sequence
  Output:   processed frame
    for each frame in video:
          normalizeImage(frame)
          resizeImage(frame, width=256, height=256)
          correctLighting(transformed_frame)
```

**A. Frame extraction:** Considering that the videos used in our system have duration of 1 second, we conducted numerous tests to determine the appropriate number of frames that can be extracted from each video; we have concluded that 10 frames are the most appropriate for our system, with each frame representing one movement. In Figure 2-5, we demonstrate dividing one video into 10 frames and converting each frame to RGB format. Subsequently, the detected human body is processed before pose detection and hand tracking. Following this, various techniques are applied to each generated frame, treating them as images.

**Figure 3-5** Ex example of dividing video into frames.

**B. Normalization:** At this level, the images are normalized so that the pixel values are between 0 and 1. This means that the initial pixel values, which range from 0 to 255 (for an 8-bit per

channel image), are converted to a range from 0 to 1. We use this process in image processing and machine learning to improve performance and model convergence. **Figure** 3-6presents a sample of the normalization technique.


**Figure 3-6** An example of normalization technique.

**C. Resizing:** The normalized image is then resized to a specific size of 256 pixels wide by 256 pixels high, which ensures that subsequent algorithms and models can operate on images with consistent dimensions, facilitating reliable and efficient processing.

**D. Lighting Correction:** We also applied the histogram normalization that can correct the lighting variations and improve the visibility of crucial features.

### 3.5.2 Sign recognition

After extracting the vector of key points, we use deep learning techniques to interpret the significance of these points. We developed four separate models for this purpose: CNN, LSTM, and a hybrid CNN-LSTM model to determine the best technique for ASL. In the following sections, we provide a detailed explanation of each component of these models.

## A. CNN model

The proposed CNN architecture that we used in our study is shown in the following figure.

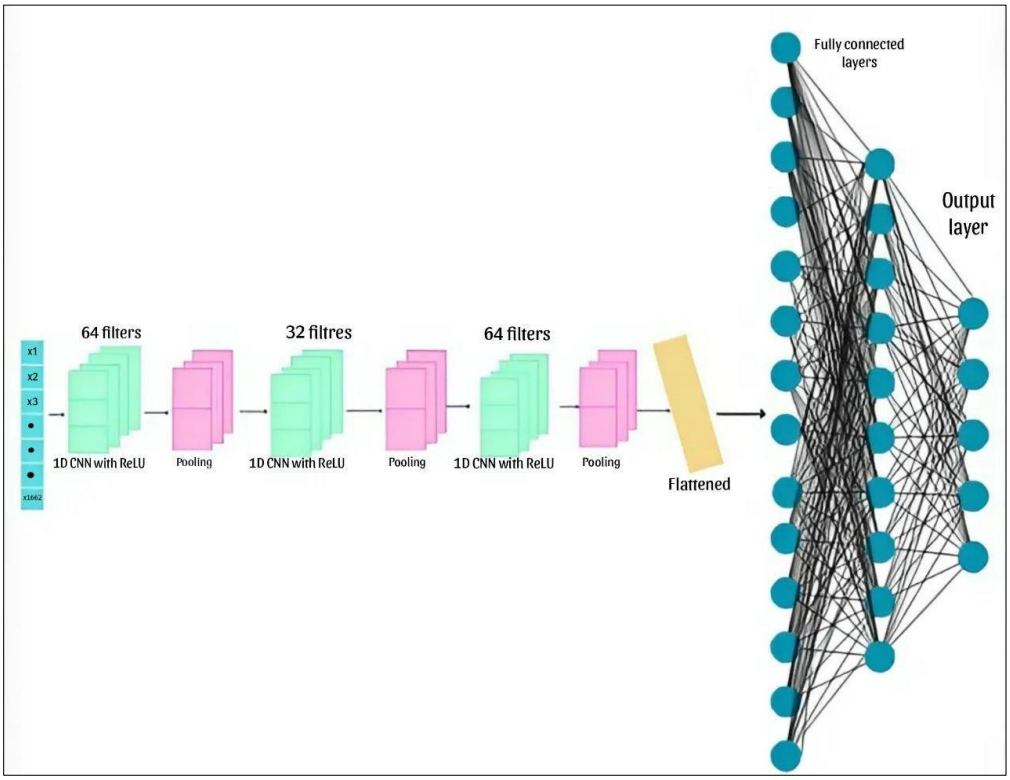



**Figure 3-7** The proposed architecture of the CNN model

In the following, the details of the provided architecture:

**a. Convolutional layers:** The model consists of three 1D convolutional layers. The first layer has 64 filters, the second layer has 32 filters, and the third layer has 64 filters, each of size 3. These filters are applied to the input vector, and a modified linear unit activation function (ReLU) is used to introduce nonlinearity to learn complex patterns. After each convolutional layer, a maximum clustering layer with a clustering size of 2 is applied. This layer reduces the size of the sequence by selecting the maximum value from every two consecutive values, which aids in feature capture and reduces computational complexity.

**b. Flatten Layer:** The Flatten layer converts the multidimensional data into a one-dimensional vector, preparing the data for connection to Dense layers.

**c. Dense Layers:** The model includes two Dense layers for classification. The first Dense layer has 64 units with ReLU activation, and the second Dense layer has 32 units with ReLU activation. The final Dense layer consists of 5 units with a softmax activation function, suitable for multiclass classification tasks such as recognizing different actions.

**d. Model Compilation:** The model is compiled using the Adam optimizer, a popular choice for weight adjustment in neural networks.

## B. LSTM model

The proposed LSTM architecture that we used in our study is shown in the following figure.

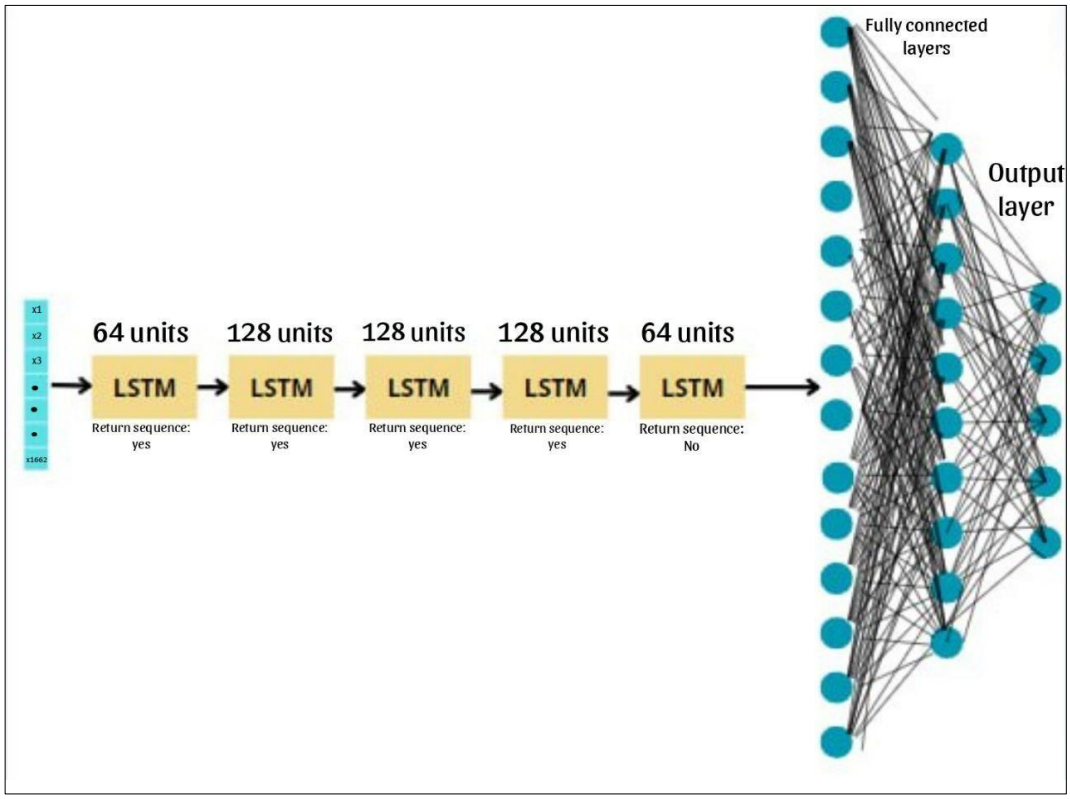



**Figure 3-8** The proposed architecture of the LSTM model

In the following, the details of the provided architecture:

**a. LSTM Layers:** The model consists of 5 LSTM layers. The first layer contains 64 units with a return sequence, the next three layers (second, third, and fourth) contain 128 units with a return sequence, and the fifth layer contains 64 units without a return sequence.

**b. Dense Layers:** The model includes two Dense layers for classification. The first Dense layer has 64 units with ReLU activation, and the second Dense layer has 32 units with ReLU activation. The final Dense layer consists of 5 units with a softmax activation function, suitable for multiclass classification tasks such as recognizing different actions.

**c. Model Compilation:** The model is compiled using the Adam optimizer, a popular choice for weight adjustment in neural networks.

### 3.5.3 Video Segmentation

This stage involves segmenting the video based on the detected tags. For example, if the video contains four tags (equivalent to four words), it will be divided into four segments of 1 second each. Each segment is treated as a single video, with the distinction that each segment represents only one word, whereas the original video may contain multiple words. Figure 3-11 shows that a single video can contain more than one word (one segment).

**A. Segment:** Each segment is considered as an input video where we can perform the same steps of detection, prediction, and output of Arabic text. By dividing the entire video into segments, with each segment being 1 second long, we ensure that each segment contains only one word.

**B. Frames:** Each segment is divided into frames, and each gesture frame will undergo the same steps of detection, prediction, and output of Arabic text on the video.

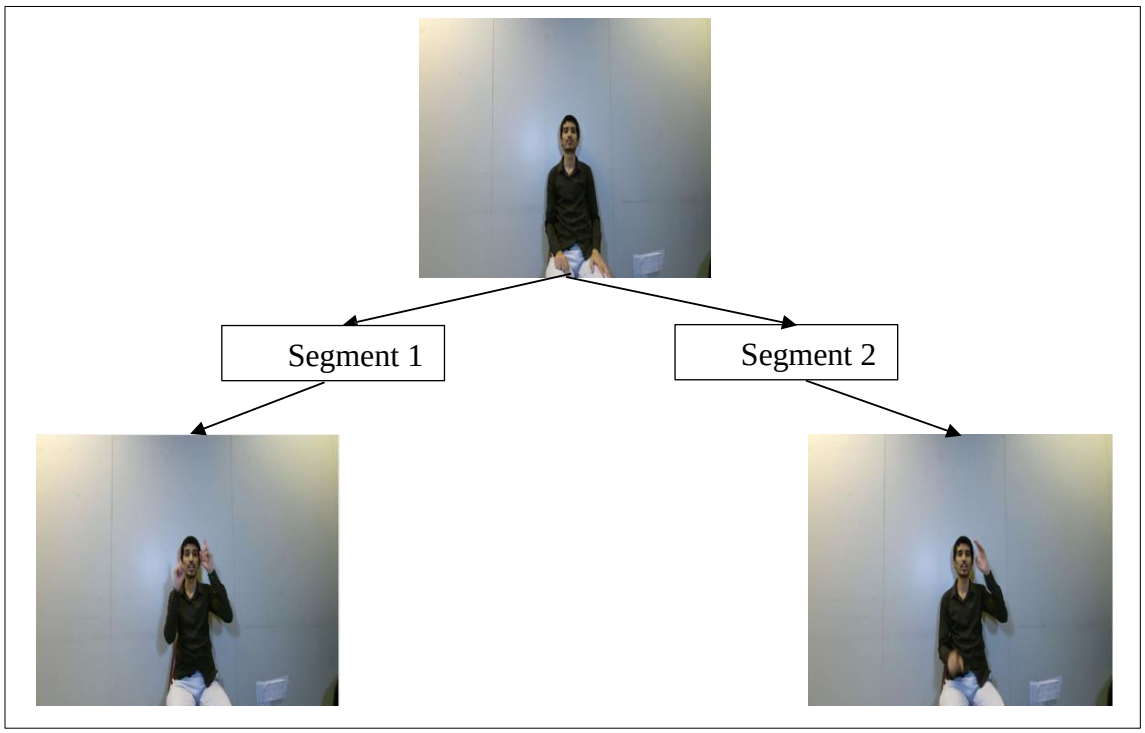



**Figure 3-9** An example' شكرا لكم' of segment generation.

## 3.6 Conclusion

In conclusion, after presenting the chapter on Algerian sign language ASL, we present the global architecture that constitutes the building blocks of our project. In this chapter we have provided a brief overview of the most important steps involved in our work, providing a clear overview of the process. In the next chapter Implementation Details we will present the architectural framework into practical solutions.

# CHAPTER-4

# IMPLEMENTATION

## 4.1 Introduction

In this chapter,we focused into the detailed conception of our system and our attention shifts towards the development environment and the libraries utilized in our system. Furthermore, we will offer insights into the essential components of our code and showcase the results we have achieved.

## 4.2 Development environment

To implement our application, we utilized a personal computer with the following specifications:

| **Model Part** | **Used Laptop** |
|---|---|
| **Processor** | AMD A8-6410 APU with AMD Radeon R5 Graphics 2.00 GHZ |
| **RAM** | 4.00 GB (3.44 GB usable) |
| **System type** | 64-bit operating system, x64 processor |
| **Edition** | Windows 10 Famille |

**Table 4-1** Characteristics of the material used.

### 4.2.1 Programming language

In this study, we used the Python language, which is detailed as follows:

**A. Python**

Python is renowned as a high-level programming language that prioritizes code readability and simplicity. Its hallmark lies in its clear and succinct syntax, enabling developers to articulate ideas using fewer lines of code compared to alternative programming languages. Python accommodates diverse programming paradigms, encompassing procedural, object- oriented, and functional programming styles. Boasting an extensive standard library and a robust ecosystem of third-party packages, Python proves versatile for a broad spectrum of applications.[26].

**B. Collab notebook**

Explore and run machine learning code with Collab Notebooks, a cloud computational environment that enables reproducible and collaborative analysis.

### 4.2.2 Libraries

In this study, we used several libraries, which are detailed as follows:

#### A. Tensorflow

TensorFlow is an open-source library that provides tools for machine learning and deep learning. It allows developers to easily create complex models, with a focus on training and inference for deep neural networks, including acquiring data, training models, making predictions, and refining future results. It is widely used in applications such as image recognition, natural language processing, and more.[27].

#### B. Keras

Keras is a high-level neural networks API, written in Python. It is designed for fast experimentation with deep neural networks, and it focuses on being user-friendly, modular, and extensible. It's particularly useful for researchers and developers in the field of deep learning.[28].

#### C. Mediapipe

MediaPipe is a cross-platform framework designed for building applied multimedia machine learning pipelines. It allows developers to create complex systems for processing time-series data such as video and audio with a focus on real-time applications, and is widely used for tasks such as hand and face detection, object tracking, and more.[29].

#### D. OpenCv

OpenCV, which stands for Open-Source Computer Vision Library, is a comprehensive open-source library that provides a multitude of computer vision and image processing algorithms. OpenCV is well-documented and supports various programming languages, including Python, which makes it accessible for a wide range of projects.[30].

## 4.3 System overview

We dedicate this section to show the interfaces of our system and the function of each of them, as our system has three different interfaces, including the frontend and the main interface, with a special focus on the main modules. The figure below shows the front end of our application.


**Figure 4-1** Home page of our system

After clic on the Start button, the basic interface shown in the image below appears

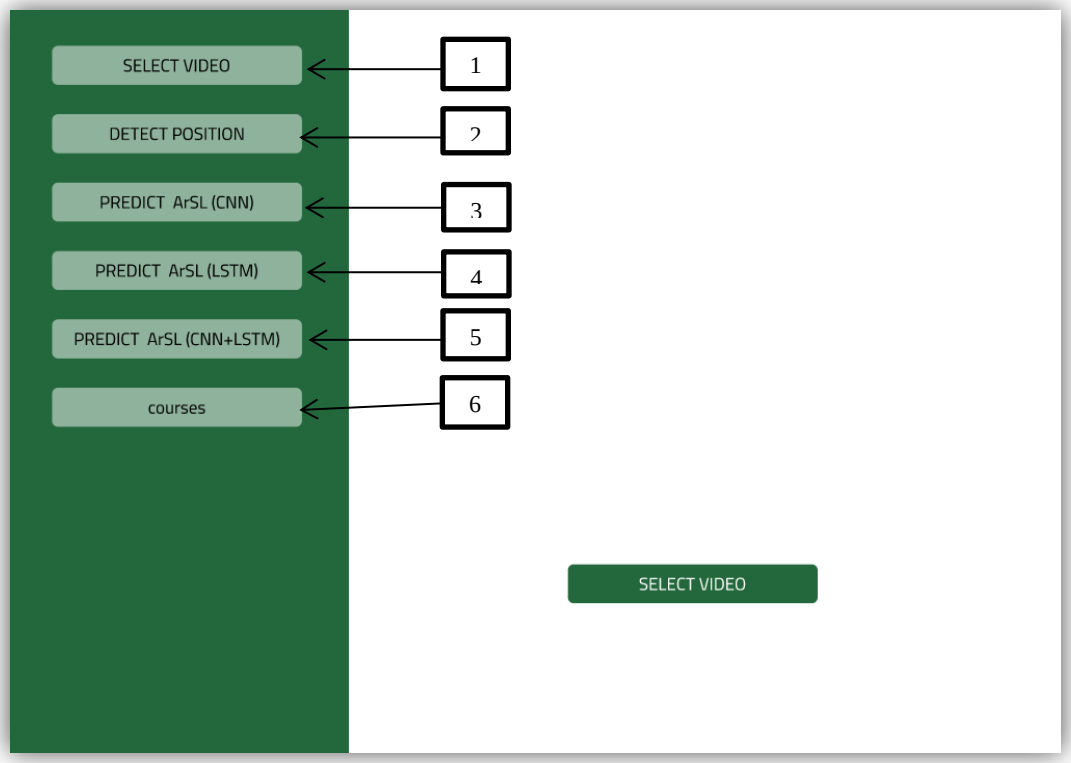



**Figure 4-2** Basic interface of our system

The main modules of our application are summarized by the following numbers:

1. The initial button enables the user to select and open video.
2. The second button is designed to perform the task of human position detection and subsequently draw styled landmarks of face hand and pose.
3. The third button is responsible for predict of Arabic sign language (ArSL) using CNN model.
4. The fourth button is utilized for predict of Arabic sign language (ArSL) using LSTM model.
5. The fifth button is responsible for predict of Arabic sign language (ArSL) using CNN&LSTM model.
6. The sixeth button is for courses

## 4.4 Usage scenario

In this section, we will provide an overview of the working principles of our system. We will describe the different stages involved in Arabic sign language recognition:

To commence the Arabic sign language recognition process, we start by opening a video file using the command "Select Video" This allows us to access the desired video and utilize it for further analysis and detection procedures. As shown in **Figure 4-3**.

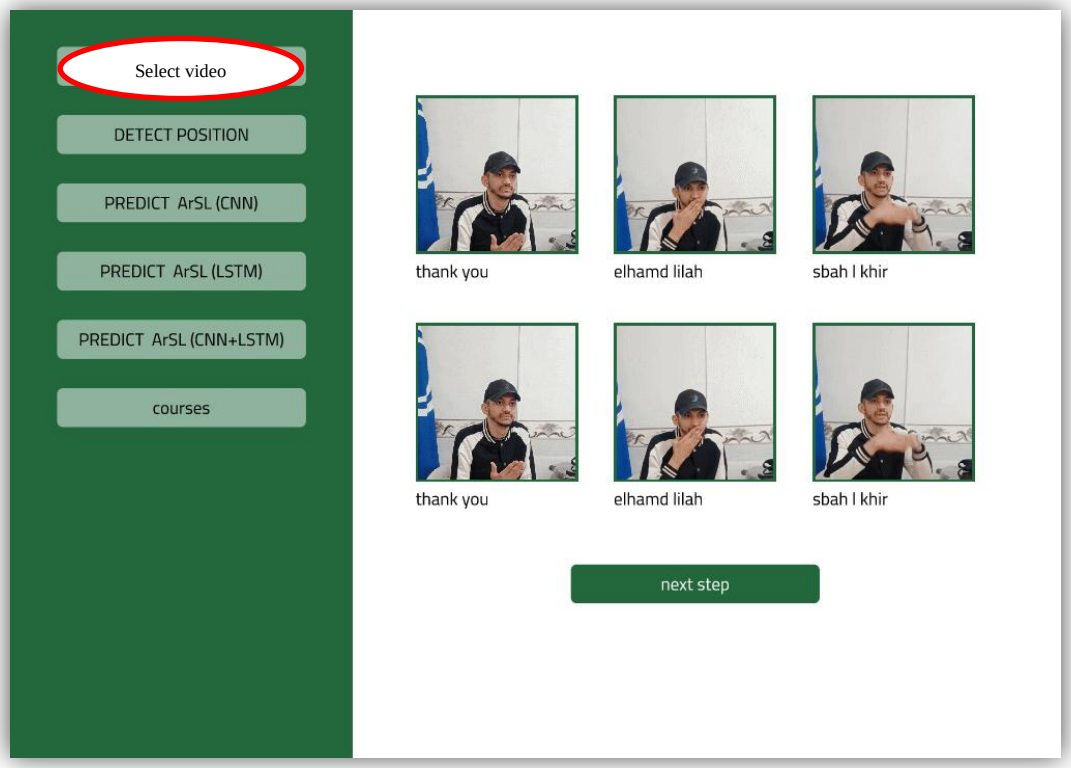



**Figure 4-3** Select video.

Once the video is successfully opened. By selecting the "detect position" button, we initiate the process of detecting the face, hand, pose of human within the opened video. The of "Mediapipe Library", which has been previously trained on a large dataset, enables accurate and efficient of human position detection by leveraging advanced object detection techniques and if the user want to saved video with detection, they clicked in button save the video saved in folder 'Detection'. The result is shown in **Figure 4-4.**

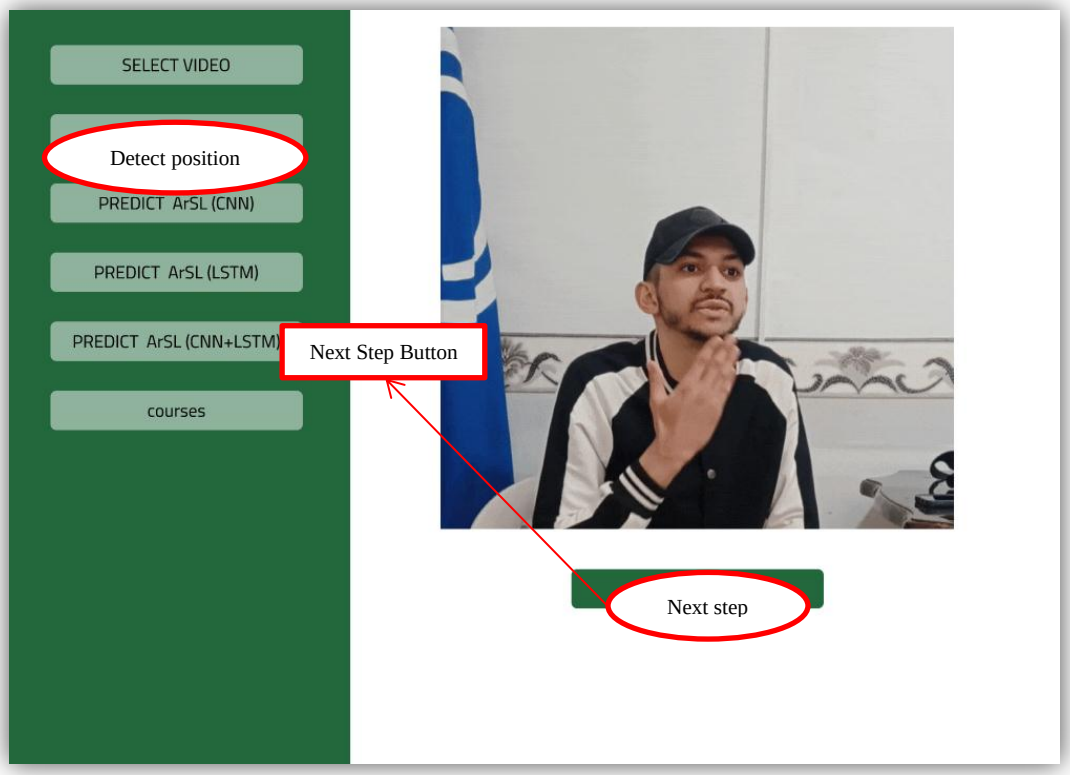



**Figure 4-4** Detect position.

The "Predict ArSL (CNN)" button predicts the labeling of Arabic words using the trained Convolutional Neural Networks (CNN) model. It splits the video into a sequence of images after that extract key points of hand and face and pose and then the model predicts the class probabilities for each key points, and then the result appears on the video in the form of sequential words as shown in Figure 4-5

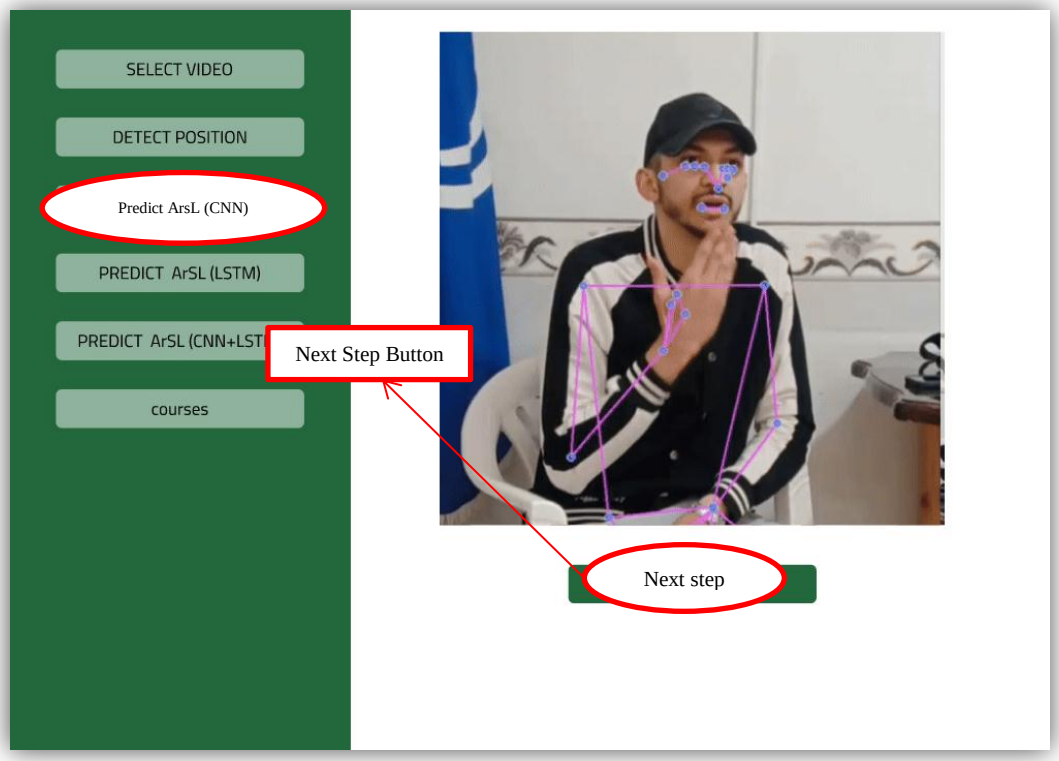



**Figure 4-5 :** Predict ArSL (CNN)

The "Predict ArSL (LSTM)" button predicts the labeling of Arabic words using the trained Long Short-Term Memory (LSTM) model. It splits the video into a sequence of images after that extract key points of hand and face and pose and then the model predicts the class probabilities for each key points, and then the result appears on the video in the form of sequential words as shown in Figure 4-6.

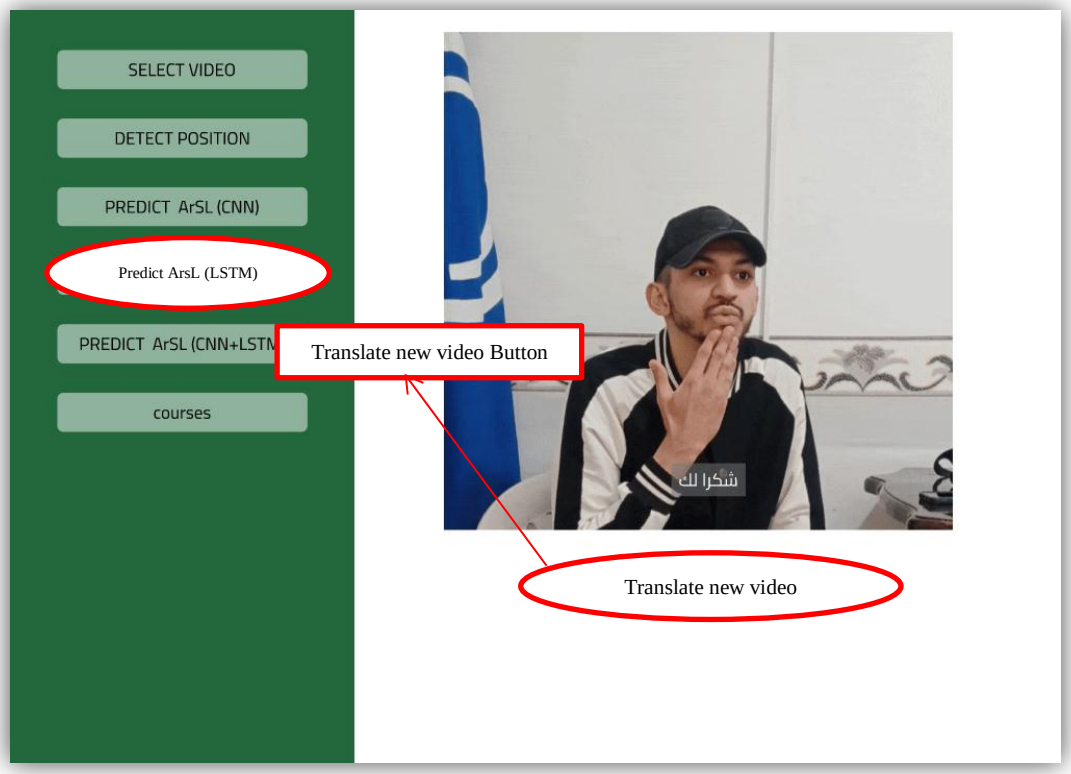



**Figure 4-6 :** Predict ArSL (LSTM)

The "Predict ArSL (CNN+LSTM)" button predicts the labeling of Arabic words using the trained hybrid between Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) model. It splits the video into a sequence of images after that extract key points of hand and face and pose and then the model predicts the class probabilities for each key points, and then the result appears on the video in the form of sequential words as shown in Figure 4-7

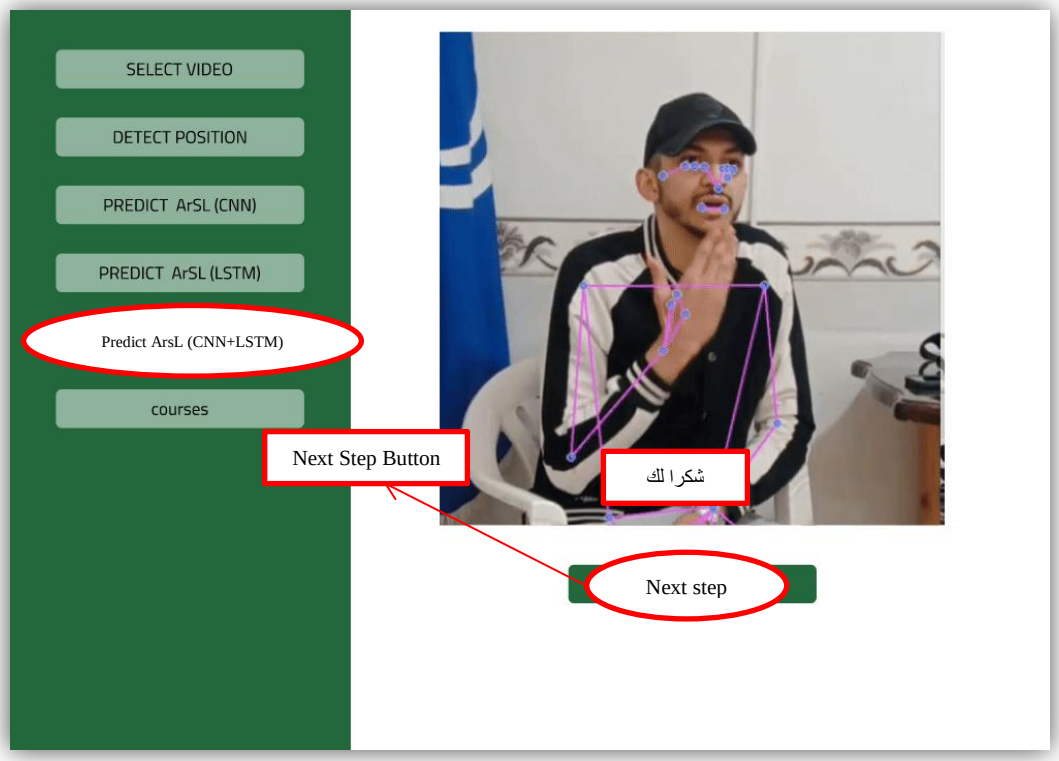



**Figure 4-7 :** Predict ArSL (CNN+LSTM)

And After clicking on the Courses button, Clicking on the Courses button will take us directly to the section for learning letters and words, and clicking on the Video button will take us directly to a video from YouTube as shown in Figure 4-8.

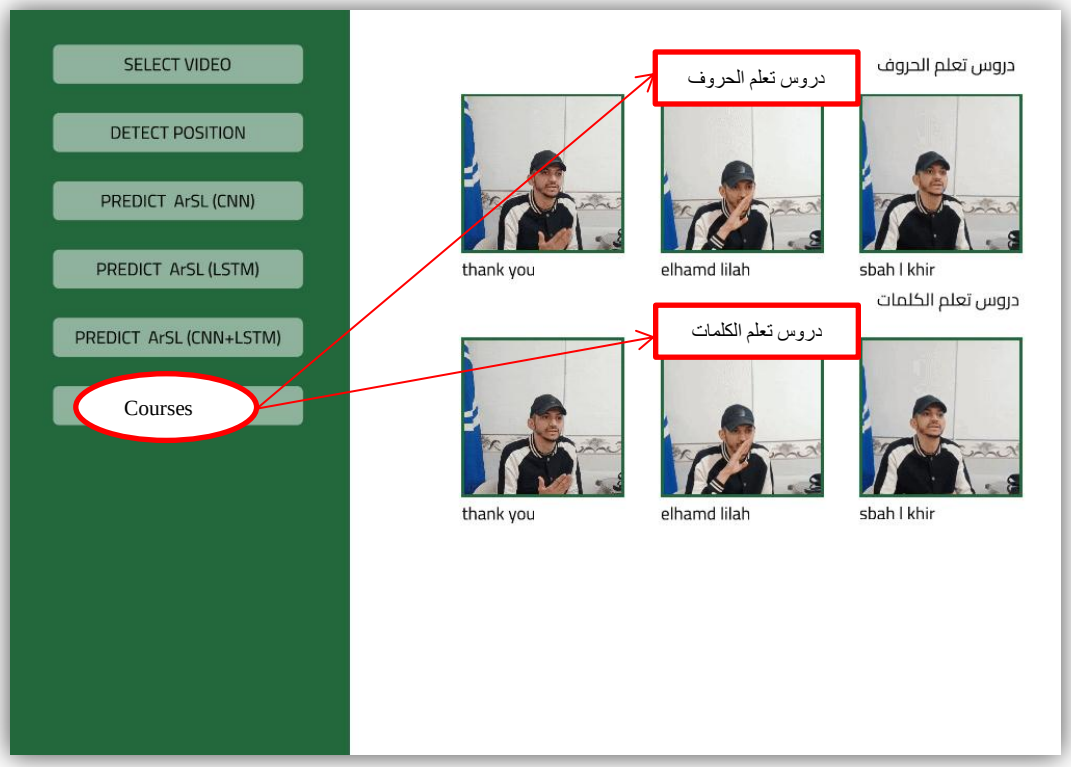



**Figure 4-8:** Courses

## 4.5 Model Performance and Analysis

In this section, we detail the results, which consist of three parts: presentation of results, analysis, and comparison of the models among themselves and with other related works that share similar features. However, due to the limited data available in the dataset we used and the computational constraints, we did not cover all words. We focused instead on the most frequent phrases that are commonly used, reflecting the nature of the dataset and our machine's processing capabilities.

### 4.5.1 Model Results

Here, we present the results of each Arabic sign language recognition model. To evaluate these models, we use metrics derived from the confusion matrix, which consists of four components:

• **True Positives (TP):** The model predicted the label and matched it correctly according to the ground truth.

• **True negatives (TN):** The model does not predict the label and is not part of the ground truth.

• **False positives (FP):** The model predicted a label, but it is not part of the ground truth.

• **False negatives (FN):** The model does not predict a label, but it is part of the ground truth. These metrics can be presented as the following:

***Accuracy:*** Accuracy measures the correctness of the results produced by a system or model. It is calculated by dividing the number of correct predictions by the total number of predictions and multiplying it by 100 to get a percentage. High accuracy indicates that a system or model is performing well in its predictions:

$$Accuracy = \frac{\text{TP+TN}}{TP+TV+FP+FN} \quad (1)$$

***Precision:*** Precision measures the ratio of correctly predicted bounding boxes for objects compared to the total predicted bounding boxes. It indicates the algorithm's ability to minimize false positives. With the following mathematical relationship:

$$Precition = \frac{\text{TP}}{TP+FP} \quad (2)$$

***Recall:*** Recall measures the number of positive class predictions made out of all the positive examples in the dataset. It answers the question "Out of all the actual positive examples, how many positive examples did the model correctly predict as positive?" We calculate it using the formula:

$$Recall = \frac{\text{TN}}{TN+FP} \quad (3)$$

***F1 score:*** is a metric used to evaluate the performance of a classification model, especially in cases where the data is unbalanced. It is the harmonic mean of precision and recall, providing a balance between the two. It is measured using the following formula:

$$F1\ score\ = 2 * \frac{\text{Precition}*\text{Recall}}{Precition+RecallN} \quad (4)$$

***Mean Average Precision (mAP):*** is used to measure the performance of computer vision models. mAP is equal to the average of the Average Precision metric across all classes in a model. You can use mAP to compare both different models on the same task and different versions of the same model. mAP is measured between 0 and 1.

***Loss:*** is a measure of how well or poorly the model's predictions match the actual results. It measures the difference between predicted and actual values. The goal in many machine learning algorithms is to minimize this loss during training, which means making the model's predictions as accurate as possible.

In our case, we have seven distinct classes as shown in the following table. As a result, we evaluate the performance of our model using the confusion matrix, where it provides information of how well our model distinguishes between each class, showing the counts of true positives, true negatives, false positives, and false negatives for each class.

## A. CNN model

The model achieved a validation loss of 0.2137, indicating that it effectively minimized the discrepancy between the predicted and actual values. This suggests that the model's predictions are generally close to the ground truth. Furthermore, the validation accuracy of 0.9523 reveals that the model correctly classified a significant proportion of the validation samples. These results indicate that the model exhibits a high level of accuracy in distinguishing between different classes or categories.

Overall, the evaluation results emphasize the efficacy and potential of the model in solving the problem at hand, validating its capability to make accurate predictions and contributing to the overall success of your research. The accuracy and loss curve are shown in Figure 4-9.

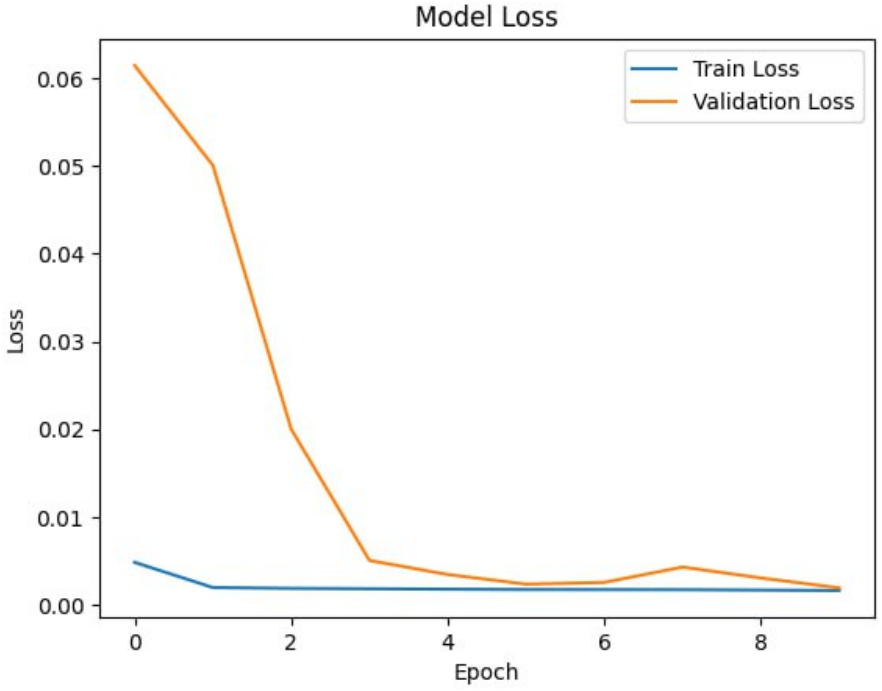



**Figure 4-9:** Accuracy and training loss Curve of CNN model.

## B. LSTM model

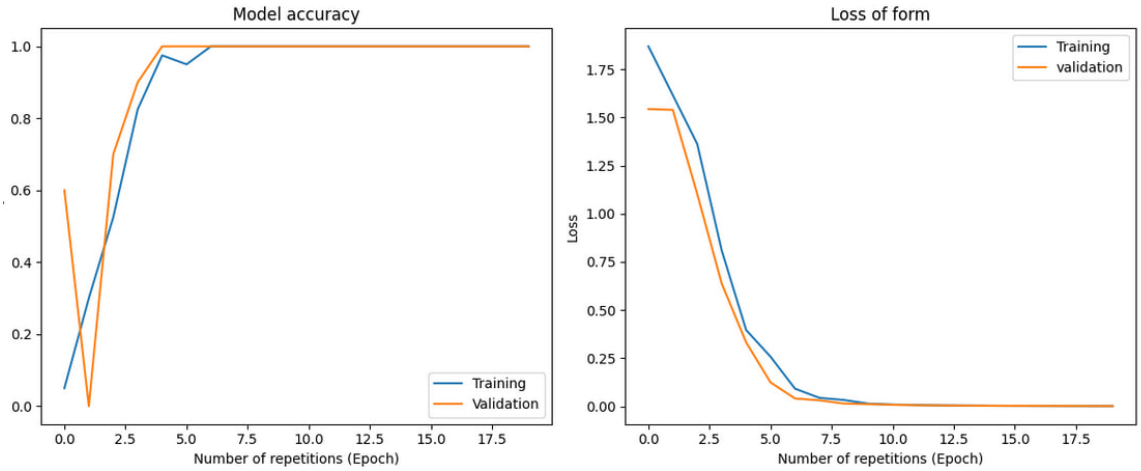
The results of the LSTM model presented and detailed in the previous chapter are shown in the **Table 4-4**

| Loss | Accuracy | Precision | Recall | F1_Score |
|---|---|---|---|---|
| 0.5862 | 0.8809 | 0.9024 | 0.8809 | 0.8915 |

**Table 4-4**: Result of LSTM model

## C. Hybrid model

As we said earlier, we wanted to optimize the results, so we combined two models CNN and LSTM into one and obtained a low validation loss of 0.0465 and an accuracy rate of 0.9666. The accuracy and loss curve are shown in :




(a)Accurracy curve (b) Training loss curve

**Figure 4-10:** Accuracy and training loss Curve of the hybrid model.

And confusion matrix of this model shown in Figure 4-11:

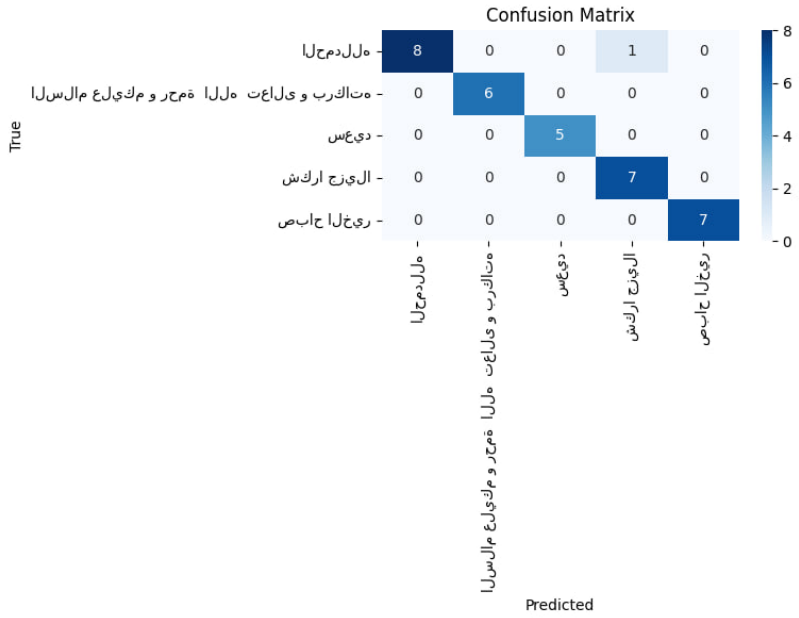



**Figure4-11:** Confusion Matrix of Hybrid model.

Confusion matrix as a table:

| index | الحمدلله | السلام عليكم و رحمة الله تعالى و بركاته | سعيد | شكرا جزيلا | صباح الخير |
|---|---|---|---|---|---|
| الحمدلله | 8 | 0 | 0 | 1 | 0 |
| السلام عليكم و رحمة الله تعالى و بركاته | 0 | 6 | 0 | 0 | 0 |
| سعيد | 0 | 0 | 5 | 0 | 0 |
| شكرا جزيلا | 0 | 0 | 0 | 7 | 0 |
| صباح الخير | 0 | 0 | 0 | 0 | 7 |

**Table 4-5:** Confusion matrix as a table

The Finall accuracy and loss curve are shown in :

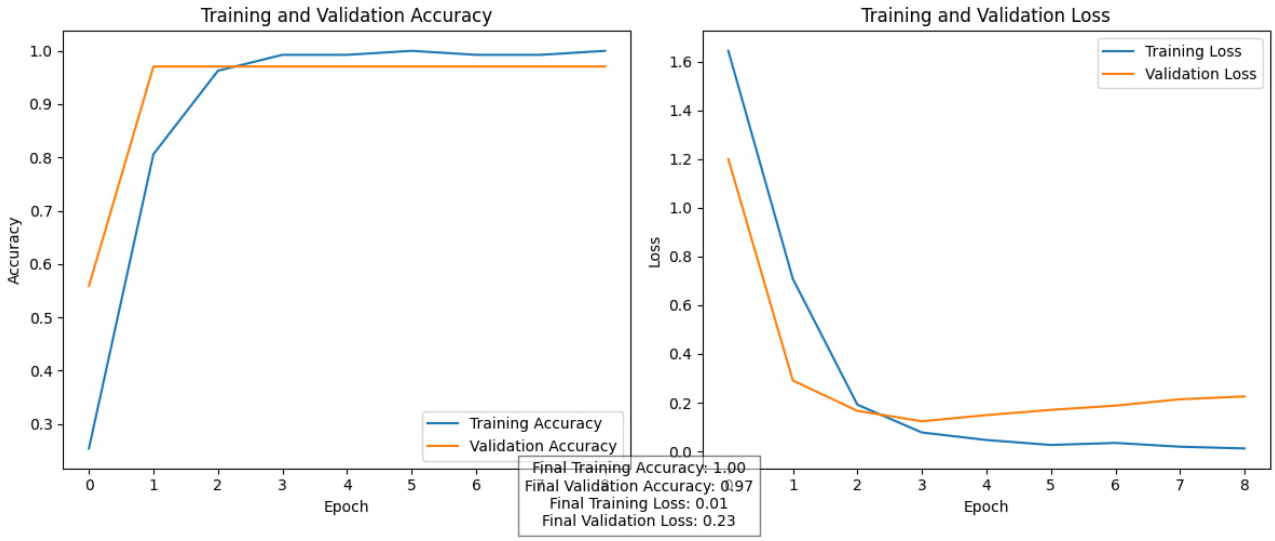



**Figure 4-12:** Accuracy and training loss Curve of the hybrid model.

## 4.5.2 Results Analysis

In this section, we provide a true comparison and analysis based on the models' results that were developed in this study which basically divided into four models. The table below represents a comparison between these four models:

| | validation loss | Precision | Recall | F1_score |
|---|---|---|---|---|
| CNN | 0.2137 | 0.9523 | 0.9523 | 0.9523 |
| LSTM | 0.5862 | 0.9024 | 0.8809 | 0.8915 |
| CNN&LSTM | 0.04653 | 0.9666 | 0.9666 | 0.9666 |

**Table 4-6** Comparison table of our system models.

From the results, we can see that our system works well in the detection and recognition phase, especially in normal situations. The high precision and recall values indicate that the system successfully recognizes Arabic sign language. This means that the system is effective in detecting and recognizing human body parts, demonstrating its robustness and reliability.

Regarding the YOLOv8 model developed for comparison purposes, it is not considered the best solution for recognizing Arabic sign language, as its accuracy is low compared to previously developed models. This may be due to the small dataset, which does not include all cases and is the same dataset used to train all models.

Based on these statistics, we can say that the hybrid model is the most effective for handling and covering the recognition task well, which was able to solve many issues and deal with many obstacles that can be characterized as follows:

1. Developed a new model based on the combination of two distinct models: CNN and LSTM.

2. Provided a real implementation of four different models, which gives more information about the models' performance in Arabic language.

3. Our system excelled in detecting and recognizing human location even under difficult conditions of light contrast and distortion.

4. Our system shows high accuracy in recognizing Arabic sign language, reliably recognizing the words that each sign represents.

5. We covered a new aspect of Arabic sign recognition where the input is a video, not an image, addressing a type of input that is very limited and covered in only a few papers.

### 4.5.3 Result discussion

We encountered difficulties comparing our work to existing studies due to the lack of similar work using the same dataset. Each study is isolated and uses hand-developed datasets specific to its research. However, we selected the closest studies that seemed similar to ours. Contrary to the results of previous studies, our proposed model (the hybrid model) achieved the best results and proved effective in recognizing sign language, achieving an error rate of 3.34%.

However, as with any system, we faced several challenges that remain, which we will outline in the following points:

1. The main problem we faced was related to the high computational cost of training deep learning models, given the large video size of over 1.31 MB per clip. Therefore, we selected a small number of words because the training process requires a large amount of RAM, which is not available on our machine. We also encountered issues with library versions when training with COLAB and using the trainer model.

Finally, we transformed this system into an Algerian Sign Language system, creating a database that includes all cases, specific to the Algerian language only, and expanding its scope to include the entire Maghreb region.

## 4.6 Conclusion

In this chapter, we provided details about our system for human position detection and Arabic sign language recognition. For the detection task, we used the Mediapipe library, which shows excellent performance in detecting the human body and its movements. In addition, we use the CNN, LSTM, LSTM&CNN models for the ArSL task. After experiments, we obtained impressive results in terms of accuracy, which shows the effectiveness of our system.

# CONCLUSION

# *CONCLUSION*

The current message in this memory allows us to comment on the automatic reconnaissance system of the signal stats in the language of the Algerian signals. This reconnaissance is based on the extraction of characteristics on the part of the image channel that represent the different signs , then their use in the training system of the learning supervise after the automatic reconnaissance of the new sign.

The second step concerns the implementation of two recognition methods, namely : the convolutional neural network (CNN) method, we achieved a rate of about 65, and long Short Term (LSTM).

Our technology is capable of automatically translating static hand gestures and facial expressions into written characters using the Algerian Sign Language alphabet.

We want to enhance our convolutional neural network (CNN) method with a sizable database of every word as future work for our system.

Additionally, we hope to develop an educational platform that instructs non-native speakers of sign language. We will be able to close the communication gap between the general population and the deaf if this research is effective.

We are also thinking about adding dynamic sign recognition to our system, which would allow us to interpret all Algerian Sign Language signs and make a variety of tasks, including employment and education, easier.

All things considered, this research has validated the viability of our sign language system, showing its potential for real-world implementation and opening the door for further developments in this area.

# REFERENCES

**References**


[1] Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3), 311-324.

[2] Kelly, D. (2010). Computational Models for the Automatic Learning and Recognition of Irish Sign Language (*Doctoral dissertation, National University of Ireland Maynooth*).

[3] Borg, M., & Camilleri, K. P. (2019, May). Sign language detection "in the wild" with recurrent neural networks. *In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1637-1641. IEEE

[4] Alani, A. A., & Cosma, G. (2021). ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. *Indonesian journal of electrical engineering and computer science*, 22.

[5] Mohamed, N., Mustafa, M. B., & Jomhari, N. (2021). A review of the hand gesture recognition system: Current progress and future directions. *ieee access*, 9, 157422-157436.

[6] Muhammad Aminur Rahaman (2018). Computer vision-based Bangla sign language recognition. *Ph. D. dissertation*.

[7] Vamplew, Peter. (1999). Recognition of Sign Language Gestures Using Neural Networks.*Neuropsychological Trends*. 1

[8] Wu, M. (2024). Gesture Recognition Based on Deep Learning: A Review. EAI Endorsed Transactions on e-Learning, 10.

[9] Farooq, U., Rahim, M. S. M., Sabir, N., Hussain, A., & Abid, A. (2021). Advances in machine translation for sign language: approaches, limitations, and challenges. *Neural Computing and Applications*, 33(21), 14357- 14399.

[10] Aliwy, A. H., & Ahmed, A. A. (2021). Development of arabic sign language dictionary using 3D avatar technologies. *Indonesian Journal of Electrical Engineering and Computer Science*, 21(1), 609-616.

[11] Wikipedia. (2019, February 18). Artificial Intelligence. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Artificial_intelligence

[12] Wikipedia. (2019, February 18). *Artificial Intelligence*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Artificial_intelligence

[13] Wikipedia. (2019, February 18). *Artificial Intelligence*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Artificial_intelligence

[14] Kanade, V. (2022, April 4). What Is Machine Learning? Definition, Types, Applications, and Trends for 2022. Spiceworks. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/

[15] Wikipedia. (2019, February 18). *Artificial Intelligence*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Artificial_intelligence

[16] Wikipedia. (2019, February 18). *Artificial Intelligence*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Artificial_intelligence

[17] Encord. (2024, April 4). *YOLO models for Object Detection Explained [Yolov8 Updated]*. Encord.com. https://encord.com/blog/yolo-object-detection-guide/

[18] YOLO Algorithm: Real-Time Object Detection from A to Z. (2015). Kili-Website. https://kili--technology-com.translate.goog/data-labeling/machine-learning/yolo-algorithm-real-time-object-detection-from-a-to-

z?_x_tr_sl=en&_x_tr_tl=fr&_x_tr_hl=fr&_x_tr_pto=rq&_x_tr_hist=true

[19] *View of The Investigation of Performance Comparison for VGG, YOLO, and DINO in Image Classification*. (2025). Drpress.org. https://drpress.org/ojs/index.php/HSET/article/view/18546/18085

[20] View of The Investigation of Performance Comparison for VGG, YOLO, and DINO in Image Classification. (2025). Drpress.org. https://drpress.org/ojs/index.php/HSET/article/view/18546/18085

[21] Great Learning. (2021, September 23). Everything you need to know about VGG16. Medium. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918

[22] GeeksforGeeks. (2020, February 26). VGG-16 | CNN model. GeeksforGeeks. https://www.geeksforgeeks.org/vgg-16-cnn-model/

[23] Great Learning. (2021, September 23). Everything you need to know about VGG16. Medium. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918

[24] Nekkaa, Foudil. Détection automatique de la main : Application à la reconnaissance de la langue des signes arabe. Magistère en Informatique. Université Abdelhamid Mehri- Constantine 2. Soutenu 2014/2015.116p

[25] Julien, Thomet. Une vire d'ensemble de la reconnaissance de gestes, Departement d'infonnatique Universite de Fribourg, vol.8.

[26] Python Software Foundation. Python. https://www.python.org/. Accessed on 10th June 2023.

[27] TensorFlow Documentation. https://www.tensorflow.org/learn?hl=fr. Accessed :29/05/2024.

[28] Keras Documentation. https://keras.io. Accessed :29/05/2024.

[29] MediaPipe – The Ultimate Guide to Video Processing. https://learnopencv.com/introduction-to-mediapipe/#What-is-MediaPipe?. Accessed :29/05/2024.

[30] What is OpenCV? – An Introduction Guide. https://pythongeeks.org/what-is-opencv/. Accessed :29/05/2024.

## Websites :

[16] L'alphabet https://vb.3dlat.net/showthread.php?t=183845. 2017. Print.

## Table references :

[1] Keylabs. (2023, December 20). YOLOv8: Object Detection Explained | Keylabs. Keylabs: Latest News and Updates. https://keylabs-ai.translate.goog/blog/under-the-hood-yolov8-architecture-explained/?_x_tr_sl=en&_x_tr_tl=fr&_x_tr_hl=fr&_x_tr_pto=rq

## Figures references :

[1] Luqman, H. (2023, January). ArabSign: A Multi-modality Dataset and Benchmark for Continuous Arabic Sign Language Recognition. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)* (pp. 1-8). IEEE.

[2] Muhammad Aminur Rahaman (2018). Computer vision-based Bangla sign language recognition. *Ph. D. dissertation*.

[3] Alani, A.A., Cosma, G., Taherkhani, A., & McGinnity, T.M. (2018). Hand gesture recognition using an adapted

**References**

convolutional neural network with data augmentation. *2018 4th International Conference on Information Management (ICIM)*, 5-12.

[4] Do, N.-T.; Kim, S.-H.; Yang, H.-J.; Lee, G.-S. Robust Hand Shape Features for Dynamic Hand Gesture Recognition Using Multi-Level Feature LSTM. Appl. Sci. 2020, 10, 6293. https://doi.org/10.3390/app10186293

[5] Farooq, U., Rahim, M. S. M., Sabir, N., Hussain, A., & Abid, A. (2021). Advances in machine translation for sign language: approaches, limitations, and challenges. *Neural Computing and Applications*, 33(21), 14357- 14399.

[6] Ismail, M. H., Dawwd, S. A., & Ali, F. H. (2021). Static hand gesture recognition of Arabic sign language by using deep CNNs. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(1), 178-188.

[7] Viswavarapu, L. K. (2018). Real-Time Finger Spelling American Sign Language Recognition Using Deep Convolutional Neural Networks *(Doctoral dissertation, University of North Texas).*

[8] Kanade, V. (2022, April 4). What Is Machine Learning? Definition, Types, Applications, and Trends for 2022. Spiceworks. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/

[9] Bansal, S. (2019, April 17). Supervised and Unsupervised learning. GeeksforGeeks. https://www.geeksforgeeks.org/supervised-unsupervised-learning/

[10] Bansal, S. (2019, April 17). Supervised and Unsupervised learning. GeeksforGeeks. https://www.geeksforgeeks.org/supervised-unsupervised-learning/

[11] https://en.wikipedia.org/wiki/File:AI_hierarchy.svg

[12] YOLOv8: A New State-of-the-Art Computer Vision Model. (n.d.). Roboflow.com. https://yolov8.com/

[13] Redmon Joseph et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016 .

[14] Great Learning. (2021, September 23). Everything you need to know about VGG16. Medium. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918

[15] Sharma, S., & Singh, S. (2022). Recognition of Indian sign language (ISL) using deep learning model. Wireless personal communications, 123(1), 671-692.

**التعرف على لغة الاشارة الجزائرية << انشاء نظام لترجمة الاشارات / الكلمات >>**


**الملخص :**

نظرا لأن الاتصال ضروري للاتصال البشري، فإن مجتمع الصم يواجه عقبات فريدة. لذلك، فإن لغة الإشارة هي أفضل بديل للتغلب على حواجز الاتصال هذه، لأنها تعتبر أكثر وسائل الاتصال فعالية، والتي تنطوي على العديد من الحركات اليدوية. ومع ذلك، غالبًا ما يساء فهم لغة الإشارة من قبل أولئك الذين ليسوا جزءًا من مجتمع الصم، مما يستلزم استخدام المترجمين الشفويين. وقد دفع هذا المجتمع إلى تطوير تقنيات لتسهيل مهام التفسير. على الرغم من التقدم المحرز في التعلم العميق، لا يزال هناك بحث محدود حول التعرف على لغة الإشارة العربية الجزائرية وترجمتها. دفعنا هذا النقص في البحث إلى التركيز بشكل خاص على تطوير الدراسات بلغة الإشارة العربية الجزائرية. تقدم هذه الأطروحة منهجيات محسنة لبناء إطار شامل لتجهيز وترجمة وتوليد لغة الإشارة العربية الجزائرية من مقاطع الفيديو المدخلة. نبدأ باستخدام مكتبة **Mediapipe** لتحديد أجزاء جسم الإنسان. بعد ذلك، من أجل التعرف على لغة الإشارة، لا سيما في اللغة العربية، استخدمنا ثلاث نماذج مميزة : الشبكات العصبية التلافيفية **(CNN)**، والذاكرة قصيرة المدى **(LSTM)** ، ونهج **CNN-LSTM**الهجين. باستخدام مجموعة بيانات **Algerian signs** ، قمنا بتكييفها للتركيز على الكلمات الفردية، وحققنا دقة. 95.23٪ لطراز **CNN** ، و 88.09٪ لطراز **LSTM** ، و 96.66٪ للطراز الهجين. تم إجراء تحليل مقارن لتقييم منهجيتنا، مما يدل على تمييز فائق بين العلامات الثابتة مقارنة بالبحث السابق.



**الكلمات الرئيسية:** لغة الاشارة العربية، **ArSL، CNN، LSTM، Hybrid CNN-LSTM، Mediapipe.**


## Recognition of the Algerian Sign Language: « the establishment of a translation system sings/words »


## Abstract:

Considering that communication is essential for human connection, the deaf community faces unique obstacles. Therefore, sign language is the best alternative for overcoming these communication barriers, as it is considered the most effective means of communication, involving many hand movements. However, sign language is often misunderstood by those not part of the deaf community, necessitating the use of interpreters. This has led the community to develop techniques to facilitate interpretation tasks. Despite progress in deep learning, there is still limited research on recognizing and translating Algerian Arabic sign language. This lack of research has prompted us to focus specifically on advancing studies in Algerian Arabic sign language. This thesis introduces improved methodologies to construct a comprehensive framework for processing, translating, and generating Algerian Arabic sign language from input videos. We begin by utilizing the Mediapipe library for identifying human body parts. Then, for sign language recognition, particularly in Arabic, we employed three distinct models: Convolutional Neural Networks **(CNN)**,

Long Short-Term Memory **(LSTM)**, and a hybrid **CNN-LSTM** approach. Using the ArabSign-A dataset, we adapted it to focus on individual words, achieving an accuracy of **95.23%** for the **CNN** model, 88.09% for the **LSTM** model, and 96.66% for the hybrid model. A comparative analysis was conducted to evaluate our methodology, demonstrating superior discrimination between static signs compared to prior research.
**Keywords:** Arabic sing language, **ArSL**, **CNN**, **LSTM**, **Hybrid CNN-LSTM**, **Mediapipe**.

## Résumé:

Considérant que la communication est essentielle à la connexion humaine, la communauté sourde fait face à des obstacles uniques. Par conséquent, la langue des signes est la meilleure alternative pour surmonter ces barrières de communication, car elle est considérée comme le moyen de communication le plus efficace, impliquant de nombreux mouvements de la main. Cependant, la langue des signes algérienne est souvent mal comprise par ceux qui ne font pas partie de la communauté sourde, ce qui nécessite l'utilisation d'interprètes. Cela a conduit la communauté à développer des techniques pour faciliter les tâches d'interprétation. Malgré les progrès de l'apprentissage profond, il existe encore peu de recherches sur la reconnaissance et la traduction de la langue des signes arabe algérienne . Ce manque de recherche nous a poussés à nous concentrer spécifiquement sur l'avancement des études en langue des signes arabe algérienne . Cette thèse présente des méthodologies améliorées pour construire un cadre complet pour le traitement, la traduction et la génération de langage gestuel arabe à partir de vidéos d'entrée. Nous commençons par utiliser la bibliothèque Mediapipe pour identifier les parties du corps humain. Ensuite, pour la reconnaissance du langage des signes, en particulier en arabe, nous avons utilisé trois modèles distincts : Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) et une approche hybride CNN-LSTM. En utilisant l'ensemble de données ArabSign-A, nous l'avons adapté pour mettre l'accent sur des mots individuels, obtenant une précision de 95,23 % pour le modèle CNN, 88,09 % pour le modèle LSTM et 96,66 % pour le modèle hybride. Une analyse comparative a été effectuée pour évaluer notre méthodologie, démontrant une discrimination supérieure entre les signes statiques par rapport aux recherches antérieures.
**Mots-clés :** Langue des signes arabe, **ArSL**, **CNN**, **LSTM**, **Hybrid CNN-LSTM**, **Mediapipe**.