CHAPTER I
Introduction

1.1. Research Background

The rapid advancements in robotics and artificial intelligence (AI) have had a transformative impact across various sectors, including education. While technological progress has opened new opportunities, significant challenges persist in achieving inclusive education, particularly in providing accessible learning tools for students with disabilities, especially those with hearing impairments. Mathematics, as one of the fundamental disciplines, often poses difficulties for students with such disabilities due to the lack of specialized teaching aids and instructional methods tailored to their needs (Linda & Muliasari, 2021). This gap is further exacerbated by the limitations of conventional teaching methods, including sign language instruction by educators, which often prove inadequate in effectiveness and coverage, particularly in situations requiring individualized attention for each student (Fitriyani, 2024) Therefore, innovative solutions are necessary to bridge this gap and ensure the accessibility and effectiveness of learning for all students.

The reliance of deaf students on teachers and the limited access to independent learning resources constitute significant challenges in inclusive education. In many cases, educators can only assist one student at a time, making the efficiency and scalability of the learning process complex (Husniati et al., 2023). Additionally, many deaf students develop individualized variations of sign language, leading to differences in their understanding of mathematical concepts based on their specific learning environments (Noviyanti & Rahmayani, 2021). Furthermore, the availability of assistive technologies remains inadequate and has not yet fully supported interactive learning or fostered autonomy in the learning process (Khasawneh, 2023). The accumulation of these challenges underscores the urgency of developing innovative solutions that facilitate more efficient and independent learning of basic arithmetic for deaf students.

In addressing these pressing challenges, artificial intelligence (AI) and machine learning have emerged as promising research fields in enhancing educational accessibility for students with disabilities. Several studies have demonstrated the effectiveness of AI-based sign language recognition systems, including those utilizing Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to interpret sign language gestures in real-time (Abdulhamied et al., 2023; Gayathri et al., 2024) The application of AI in sign language identification offers scalable solutions for developing high-accuracy learning tools capable of recognizing and understanding mathematical gestures (Saleem et al., 2023).

In line with these advancements, robotic hand technology has emerged as an innovative solution in sign language communication. Research has shown that robotic hands capable of demonstrating sign language letters hold significant potential in supporting the learning process of deaf students (Angga et al., 2019). Furthermore, the development of robotic systems equipped with YOLO Keypoint Detection presents an opportunity for this technology to effectively identify and interpret hand gestures representing numbers (0–9), arithmetic operations (+, -, ×, ÷), initiation symbols, and other gestures. This method provides real-time feedback to students while solving arithmetic problems, enhancing interactivity and efficiency in the learning process (Fadlilah et al., 2022). Additionally, YOLO-based motion recognition technology has demonstrated superior accuracy and speed compared to conventional computer vision methods, making it an ideal solution for real-time educational applications (Ranasinghe et al., 2022).

Building upon this potential, the development of an AI-integrated robotic hand system based on YOLO Keypoint Detection is proposed. This system is designed to recognize 16 classes, including 10 numerical digits, four arithmetic operators, one initiation symbol, and one undefined symbol, thereby providing optimal support for deaf students in mathematics education. To ensure the model's robustness and generalization capabilities, the dataset will include hand gesture images from various angles, distances, and lighting conditions, improving the system's accuracy and adaptability across different learning environments. Similar implementations in machine learning-based educational platforms have shown significant success in improving the academic performance of deaf students, particularly in bridging the gap between sign language and conventional learning approaches (Raji et al., 2023). thus, the proposed model has the potential to revolutionize accessibility and engagement for deaf and hard-of-hearing students in mathematics education through an interactive, autonomous, and scalable solution.

Building upon these established findings by (Raji et al., 2023), this research focuses purely on filling the identified technological gap. Rather than re-testing the educational effectiveness of assistive tools on students which has already been proven beneficial by previous pedagogical studies this study concentrates entirely on the technical engineering of an advanced, real-time interactive robotic prototype. This focus ensures that the architectural advantages of the latest deep learning models can be effectively combined with accessible edge hardware to establish a solid foundation for future automated teaching media.

Because the primary objective is to resolve the engineering challenge of building this real-time system, certain technical boundaries have been established. While this model is designed to assist the learning process, it functions as an indirect assistive tool. This research focuses strictly on the technical development of the hardware and software pipeline and does not test the educational effectiveness of the tool directly on human subjects. Furthermore, the integrated system and robotic hand developed here represent only a prototype development phase, serving as an initial proof of concept to demonstrate high-speed local inference on resource-constrained devices.

Despite these structural boundaries, the development of this prototype yields significant practical value. Beyond its primary application in inclusive special education programs, this interactive prototype can also be utilized by individuals and students without special needs as an engaging, tactile medium to learn basic arithmetic and sign language simultaneously. Consequently, the system serves a dual purpose: providing an open-source technical template for advanced edge-based gesture recognition while acting as an interactive bridge that fosters inclusive communication skills among a broader student demographic.

The selection of the YOLO (You Only Look Once) architecture is justified by its recognized speed and high accuracy in real-time object detection, making it highly suitable for the precise keypoint identification required in sign language recognition (Terven et al., 2023) Quantitative studies on recent YOLO versions demonstrate this capability. For instance, an optimized YOLOv4-CSP model developed for Turkish Sign Language achieved a mean Average Precision (mAP) of 99.49% with an inference time of just 9.5 ms (Alaftekin et al., 2024). Similarly, a proposed framework using YOLOv11 for American Sign Language reached a 98.1% mAP, highlighting its suitability for real-time applications on mobile and edge devices due to its efficient computational architecture (Elgohr et al., n.d.). This performance reinforces YOLO's superiority over competing models, such as MediaPipe and CVZone (Casas et al., 2023), as its approach often results in lower latency, which is crucial for high-speed applications (Lin et al., 2020). However, achieving this level of performance necessitates robust training strategies, including the use of diverse datasets that account for varied environmental conditions to ensure the model's generalization and implementation applicability (Muñoz Galindez et al., 2025).

This performance stands in favorable contrast to other methodologies. While some specialized deep learning models, such as Transformer-based architectures, have achieved exceptionally high accuracies of up to 99.74% on specific datasets, like Arabic Sign Language, YOLO's strength lies in its robustness (Elgohr et al., n.d.). Many standard CNNs, such as MobileNetV2, which achieved 97% accuracy, are highly dependent on extensive image preprocessing (e.g., background elimination and color space conversion) to perform optimally (Perdana et al., 2021). Advanced YOLO architectures, however, can operate effectively without uniform backgrounds or complex preprocessing, making them more adaptable to varied, real-world conditions (Alaftekin et al., 2024).

Furthermore, YOLO's deep learning approach significantly outperforms traditional machine learning classifiers. A system using Local Directional Pattern

(LDP) with a K-Nearest Neighbor (K-NN) classifier, for example, achieved a peak accuracy of only 88.45% (Ilmi & Suryoprayogo, 2022). The alternative of using sensor-based systems, such as data gloves, also presents drawbacks; At the same time, they can achieve high offline accuracy, but they are often impractical for daily use, and their real-time performance can be significantly lower, with one study reporting a drop to an average of 75% accuracy during live testing (Wungow et al., 2022; Alaftekin et al., 2024). Therefore, the evidence strongly supports the selection of a vision-based YOLO architecture, as it provides a superior and well-documented combination of high accuracy, speed, and environmental flexibility for the proposed real-time system. In summary, based on these considerations, the development of an AI-based robotic hand system with YOLOv11 architecture to recognize 16 mathematical sign language classes including 10 numerical digits, four arithmetic operators, an initiation symbol, and an undefined symbol is projected to significantly improve accessibility in learning basic arithmetic skills for deaf students. This approach enhances learning efficiency and provides a more interactive and engaging educational experience.

The success of this system's implementation depends significantly on the advantages inherent in YOLOv11's architecture, the optimization of model training strategies, and the availability of representative datasets, as proposed by Muñoz Galindez et al. (2025), Comparative performance analysis with other key point detection models ensures system effectiveness. With strong evaluation metrics, including F1-score, precision, recall, and mean Average Precision (mAP), this system is expected to achieve real-time hand gesture recognition with high accuracy, ultimately contributing to the improved effectiveness and accessibility of arithmetic learning for deaf students (Casas et al., 2023).

1.2. Problem Identification

Based on the background discussion above, the following problem identification has been established:

a. A dataset of simple arithmetic operations is required to serve as training data for the YOLO Keypoint Detection model.
b. A computer vision-based model is needed to accurately recognize, interact with, and respond to simple arithmetic operations using a robotic hand as the communication medium.
c. A method is needed to effectively implement and integrate the YOLO Keypoint Detection model onto a portable, low-power hardware platform, specifically a Raspberry Pi.

1.3. Research Questions

Based on the problem identification above, the research problem can be formulated as follows:

a. How is the detection procedure of simple YOLO Keypoint Detection based recognition arithmetic?
b. What is the performance of the sign language recognition model as the core of the interactive robotic hand based on YOLO Keypoint Detection?
c. How is the implementation of YOLO Keypoint Detection model a interactive Raspberry Pi based robotic hand

1.4. Research Objectives

Based on the identified problems, this research aims to:

a. Develop a system with YOLO Keypoint Detection Model that able to recognize and calculate a simple arithmetic sign language.
b. To evaluate the performance of the integrated system, specifically testing the YOLO recognition model as the core component of the interactive robotic hand and how it perform on portable Raspberry Pi.
c. Implement the developed YOLO model on a physical Raspberry Pi robotic hand, demonstrating its function as an interactive learning tool based.

1.5. Research Limitations

The scope and limitations of this research are as follows:

a. The study will focus on developing a model capable of recognizing numerical digits and arithmetic operation symbols, implementing arithmetic systems, and integrating them into a robotic hand to assist student learning.
b. The model will be designed to recognize numerical digits from 0 to 9 and four basic arithmetic symbols. This limitation to single-digit, static hand gestures is crucial for maintaining manageable model complexity. Recognizing multi-digit numbers would require interpreting dynamic gestures, which involve motion over time and necessitate more complex architectures (such as hybrid CNN-LSTM models) to process temporal sequences, significantly increasing the model's size and computational requirements.
c. The system will accept the last numerical or arithmetic operation symbol input by the user before changing the shape of their hand.
d. The robotic hand constructed will be restricted to three degrees of freedom (3-DOF) for each finger, with no wrist movement. In the context of this

research, it functions solely as a visualization tool or “monitor” to demonstrate the output of the sign language recognition model. Its mechanical accuracy and potential movement errors are considered outside the primary scope of this evaluation.

e. The model will be optimized for the Raspberry Pi 4B and its specific camera module. While the use of multiple camera types during data collection aims to improve generalizability, the model's performance on future hardware iterations (e.g., newer Raspberry Pi models or different camera sensors) is not guaranteed and represents an avenue for future work.

1.6. Research Significances

The expected benefits of this research are as follows:

a. Theoretical Benefits :

   i. This research aims to serve as an implementation of knowledge acquired throughout the educational process.

b. Practical Benefits :

   i. Contributes to creating a dataset containing sign language images for simple mathematical operations.

   ii. Advances in artificial intelligence development, particularly in computer vision-based YOLO Keypoint Detection.

   iii. Serves as an initial stage in developing a fully functional robot capable of more complex communication and interaction.