Arab American University

Faculty of Graduate Studies

Artificial Intelligence System for Arabic Sign Language Recognition to Enhance Education for Deaf and Hard-of-Hearing Students.

By

Fidaa Said Fares Khandaqji

Supervisor

Dr. Huthaifa AL- Ashqar

This thesis was submitted in partial fulfilment of the requirements for the Master`s degree in Artificial Intelligence

December/ 2025

Thesis Approval

Artificial Intelligence System for Arabic Sign Language Recognition to Enhance Education for Deaf and Hard-of-Hearing Students.

By

Fidaa Said Fares Khandaqji

This thesis was defended successfully on 27.12.2025 and approved by:

Committee members	Signature
1. Dr. Huthaifa I. Ashqar: Main Supervisor	
2. Dr. Abdelrahem Atawnih: Member of Supervision Committee	
3. Dr. Mohamed Khalil: Member of Supervision Committee	

Declaration

I declare that the content of this master's thesis is the product of my own independent research and effort. All ideas, data, and materials taken from other sources are fully recognized and appropriately cited. No material prepared by another person has been included without clear acknowledgement.

Every quotation, figure, and dataset used in this work has been properly identified and referenced. To the best of my knowledge, this dissertation has not been previously presented, in whole or in part, for any academic qualification at any university or institution.

Student Name: Fidaa Said Fares Khandaqji

Student ID: 202226026

Signature: Fidaa Khandaqii

Date: 13.1.2026

Dedication

I dedicate this humble research to my family, whose blessings have shaped my life, nurturing me with care, wisdom, and the true spirit of generosity. May your lives be filled with peace and grace.

To my beloved, my children, this work stands as a testament to perseverance, faith, and the inspiration you will always find in the pursuit of knowledge and purpose.

I tenderly dedicate this research to the pure souls of my good deceased father and brothers, praying that God grants them eternal mercy and peace.

With heartfelt appreciation, I extend my deepest gratitude to my supervisor, Dr. Huthaifa AI-Ashqar, whose guidance, patience, and encouragement have been a light throughout this journey.

Finally, I would like to dedicate this accomplishment to my dear friends and everyone who supported, encouraged, and believed in me along the way.

Acknowledgment

I want to express my sincere thanks to and my appreciation of the Palestinian Ministry of Education – General Administration of Special Education and Institutions for the deaf and hard of hearing for their generous cooperation and valuable assistance in providing the data and information necessary to complete this thesis, which had a significant impact on the success of this research work and the achievement of its objectives.

Special thanks are extended to Mr. Khalil Alawneh, an expert in Palestinian Sign Language, for his professional guidance and valuable contributions to the practical aspects of Palestinian Sign Language.

My deepest gratitude also goes to my university, the Arab American University, for its continuous academic and administrative support.

Finally, I would like to thank all my professors who supported and encouraged me throughout the preparation of this thesis. I appreciate your help and kindness.

Abstract

This study aims to examine the development and effectiveness of the “SignPulse” system, an intelligent educational system designed to support deaf and hard-of-hearing students in learning STEM education using Palestinian Sign language(PSL). The research adopts a mixed-method approach, combining qualitative and quantitative analyses to evaluate the system's impact on student engagement, comprehension, and interaction.

Data were collected through a structured video dataset for PSL, along with usability questionnaires completed by experts in sign language, educational technology, and STEM education. This dataset enabled the training and evaluation of a hybrid Siamese vision transformer (Siamese-ViT) model for accurate sign recognition learning, while the system's interactive component, powered by GPT+RAG, provided personalized explanations, contextual questions, and immediate feedback to enhance the learning experience.

The findings suggest that integrating AI technology significantly enhances learning accessibility, engagement, and comprehension for students with hearing disabilities. Moreover, the study highlights the pivotal role of specialized guidance and culturally appropriate sign language resources in maximizing the system's effectiveness. It also discusses challenges related to technical limitations, user adaptation, and integration into classroom practices.

The study concluded that AI-driven sign language systems, such as “ SignPulse,” have significant potential to improve inclusive education for students with hearing disabilities, support the implementation of adaptive learning strategies, and contribute to providing equitable educational opportunities in Palestine.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbr	**abbreviation(s), abbreviated**
DHH	Deaf and Hard of Hearing
KPI’s	Key Performance Indicators.
PSL	Palestinian Sign Language
SL	Sign Language
AI	Artificial intelligence
DL	Deep learning
STEM	Science, Technology, Engineering, and Mathematics
ViT	Vision Transformer CNN
CNN	Convolutional Neural Network
CNN-RNN	Convolutional Neural Network-Recurrent Neural Network Hybrid
BERT	Bidirectional encoder representations for transformers
CSL	Chinese Sign Language

BSL	British Sign Language
ISL	Indian Sign Language
LSTM	Long short-term memory
NMT	Neural Machine Translation
RGB	Red, Green, Blue
LLMs	Large language models
RAG	Retrieval Augmented Generation
VAE	Variational Autoencoder

Chapter One: Introduction

1.1 Introduction

Education is one of the fundamental pillars of societal development and progress. It plays a crucial role in shaping individuals and enhancing their intellectual and cognitive abilities, which leads to improved quality of life and enhanced economic and social stability. Education is not merely about acquiring skills; it also encompasses the development of critical thinking, the ability to solve problems, and fostering innovation and creativity. It is also an effective tool for achieving social justice and bridging economic gaps, as it opens equal opportunities for everyone, regardless of their social or economic background. With the rapid advancement of technology, education has become even more essential in preparing individuals for the digital job labor market. The modern economy requires advanced skills in artificial intelligence (AI), deep learning (DL), programming, and data science (DS). Therefore, adopting innovative educational strategies that leverage technology and AI is crucial to creating more interactive and inclusive learning environments in the future.

The education system in Palestine faces challenges due to political and economic conditions that have impacted its infrastructure and management, over the decades, the Palestinian education system has been under the control of various authorities, affecting its stability and development, despite the transfer of responsibility for education to the Palestinian Authority following the Oslo according in 1993, obstacles remain such as restrictions on school construction and the obstruction of students and teachers from accessing educational institutions, these barriers constitute a violation of the right to education as guaranteed by international conventions, in addition, preschool education is subject to the supervision of the ministry and higher of education, which grants licenses and oversees necessary for the standards essential for the operation of kindergartens, which are primarily run by privet sector institutions and non-governmental organizations, with teachers required to have certifications in early childhood education and adhere to standardized curricula to ensure quality education. (Hasan, & Buheji., 2024).

Despite these challenges, the Palestinian education system has remarkable resilience and adaptability. Educational institutions seek to develop more

comprehensive curricula and improve the quality of their use of technology and innovation. However, there remains a significant gap in providing an inclusive learning environment for people with disabilities, particularly the deaf and hard of hearing (DHH), who face difficulties accessing appropriate educational content. This highlights the need for AI and DL technology to bridge this gap. Developing systems capable of automatically recognizing PSL and integrating it into interactive educational tools will contribute to enhancing learning opportunities for these students and making the educational process more inclusive and efficient.

Hearing loss is a growing global problem affecting millions of individuals across age groups and geographic groups. As advances in healthcare and assistive technology continue, addressing the challenges faced by individuals who are deaf or hard of hearing remains a pressing priority. The increasing prevalence of hearing loss necessitates the development of comprehensive communication and education solutions to ensure equal opportunities for individuals affected by it. The World Health Organization (WHO (2021) repeated that more than 430 million people worldwide suffer from disabling hearing loss. It is projected that by 2050, approximately 2.5 billion people may be affected by this loss to some degree, and around 700 million of them are likely to require hearing care and rehabilitation services (WHO (2021).

The Arab world has some of the highest rates of deaf and hard-of-hearing individuals, with some countries exceeding 2% of the population (Unescwa, 2018). However, due to a lack of awareness, society often overlooks their needs, and to address this, researchers have started developing new technologies to help the deaf communicate more easily (Alnahhas et al., 2020).

According to the 2017 census, people with disabilities constitute 5.8% of the total population of Palestine, equivalent to 255,228 individuals out of 4.78 million people, of these,1.6% approximately 76,480 individuals suffer from hearing impairment, with 46,080 of them residing in the West Bank and 30,400 in Gaza, this percentage has witnessed an increasing increase, highlighting the urgent need to enhance accessibility and support provided to the deaf and hard of the hearing group, as deaf individuals in Palestine face significant challenges in various aspects of life, including education, employment, access to media, social interaction, and interpretation services, although education is a basic right for all, deaf and hard of

hearing students still face significant obstacles that hinder their academic success and full participation in the educational process, these challenge absence of comprehensive educational policies, weak resources, and shortage of qualified teachers in PSL, the lack of specialized curricula limited access to specialized schools, and the absence of adequate technological support make it more difficult for deaf students to obtain a high-Quilty education, addressing these issues is crucial to ensuring that equal educational opportunities and enable them to integrate more effectively into society (Alawneh,& Abdel-Fattah., 2021).

Sign language (SL) is a distinctive mode of human communication that holds an essential role in interaction and expression for many members of society, a language system that provides an integrated visual gestural means of expression, enabling users to communicate their thoughts, feelings, and sensations through structured hand shapes, movements, and facial expressions, rather than relying on spoken words or sounds (Rastgoo et al., 2021). Language can be expressed and understood through both spoken words and manual signs, demonstrating its adaptability across different modalities. While the legitimacy of sign languages has been debated, they are now recognized as fully developed linguistic systems. This recognition has contributed to a deeper understanding of language structure and has led to a change in our perception of human communication (Goldin-Meadow, & Brentari., 2017).

Currently, more than 300 different sign languages are used worldwide, each with its own unique grammar and syntax that distinguishes it from other languages. Such as include American Sign Language (ASL), British Sign Language (BSL), Arabic Sign Language (ArSL), and Chinese Sign Language (CSL) (Miah et al., 2024) . Recently, sign language recognition (SLR) has emerged as a key tool for improving human-computer interaction (HCI) and supporting communication available to people with hearing difficulties (Palanisamy et al., 2024).

Palestine Sign Language (PSL) is a national sign language used in the Levant region and is the primary means of communication among members of the deaf community in Palestine, although Arabic is the native language and English is the second language, PSL occupies an important position as a third language for people with hearing disabilities. A comprehensive study was conducted on PSL, and four other Arabic sign language (ASL), Kuwait sign language (KSL), Libyan sign language

(LSL), and AL-Saeed Bedouin sign language (ABSL) analysing key linguistic features such as hand shape, movement, position, and palm orientation, the result showed that these languages are not merely dialects, but rather independent linguistic systems, despite their shared community with spoken Arabic, the study also showed that PSL shares 58% similarity with indicating strong mutual influence between them, however, PSL remains largely undocumented, with only a few basic dictionaries available as references (Abdel-Fattah, 2020).

Deaf and hard-of-hearing (DHH) students face multiple challenges in the digital learning environment, where limited technical resources and weak digital curricula hinder effective learning. Studies indicate that the most prominent obstacles for teachers in this group include a lack of digital content and the weak recruitment of specialized educational frameworks, which negatively impact the quality of education (السالم والزهراني، 2021)Project-based learning is an effective educational method that fosters critical thinking and problem-solving among DHH students. However, resources and a lack of teacher training are significant barriers to its effective implementation; therefore, AI can help enhance engagement and improve the learning experience for these students (عبد العزيز الخضير, 2022).

Artificial intelligence (AI) has become one of the most influential technological advancements across multiple fields, particularly in education, computer vision, and NLP. AI has emerged as an important tool in digital education, as it customizes educational content to meet the needs of each learner, especially those with special needs, including the deaf and hard of hearing(Abulibdeh, 2025) . AI-powered technologies also provide educational experiences adapted to handle learning subjects and goals of each student (Lahby et al., 2024), which is of great importance and benefit to deaf and hard-of-hearing learners (Kamalov et al., 2023). Applying AI tools and technologies in education makes learning environments more personalized and adaptive, which meets the needs of people with disabilities (Owoc et al., 2019).

Furthermore, AI education should not be isolated or limited to specific fields but should be integrated into, linking different curricula and the educational environment to which students belong (Aliabadi et al., 2023). Despite significant advances in AI and its role in developing more efficient educational systems, challenges remain related to providing equal opportunities for all groups. This calls for the

development of educational strategies that ensure the integration of deaf and hard-of-hearing students into digital educational environments and their full benefits from these technologies. After reviewing the most important challenges facing the deaf in Palestine, it is noted that there is a lack of research related to PSL, indicating an urgent need for in-depth studies in this field. Studies have demonstrated that AI and deep learning (DL) have made substantial contributions to the advancement of intelligent systems, particularly in the fields of computer vision and natural language processing. These technologies can improve communication for deaf and hard-of-hearing people by developing real-time sign language interpretation systems, enhancing their integration into society.

Artificial intelligence (AI) refers to efforts to develop machines capable of acting intelligently; intelligence here is defined as the ability to interact appropriately and thoughtfully with the surrounding environment, taking future expectations into account when making decisions (Chassignol et al., 2018). Educational institutions have adopted AI extensively to provide various services to students in different ways. AI has been effectively integrated into enhancing the efficiency of administrative and educational processes, and studies show that AI has simplified administrative tasks, such as assessment marking and providing feedback, through automated systems and digital platforms (Chen et al., 2020).

Natural language processing (NLP) is one of the most popular AI methodologies in student opinion mining (Estrada et al., 2020). This technology plays a crucial role in understanding and interpreting the feedback or opinions expressed by end-users. Most organizations around the world devote time and effort to analyzing this feedback to understand the needs and opinions of their users (Shaiket al., 2022). Computer Vision enhances teaching methods by enabling teachers to adjust their strategies in real-time, helping improve students' academic outcomes and support teacher-student relationships, especially for those with learning difficulties (Sophokleous et al., 2021).

In computer vision, AI improves image and video recognition, object detection, and gesture analysis, enabling advanced applications such as autonomous systems, medical imaging, and assistive technologies for people with disabilities. Additionally, NLP has revolutionized human-machine interaction, enabling AI systems to

understand, interpret, and produce human language. This has led to advancements in machine translation, sentiment analysis, and interactive language models such as ChatGPT.

DHH students face significant challenges in communication and understanding academic content, as traditional educational curricula rely primarily on text and audio materials. This problem is particularly prominent in STEM science, technology, engineering, and mathematics. Complex explanations require visual and interactive support to enhance understanding. Despite advances in sign language recognition technologies, research has primarily focused on American Sign Language (ASL) and British Sign Language (BSL), resulting in significant neglect of PSL. This has led to the study of PSL being largely neglected. In the Palestinian context, AI and deep learning can contribute to a radical transformation in the field of education and communication for the hearing-impaired, by providing innovative solutions that consider the linguistic and cultural specificities of this group.

1.2 Statement of Problem

Despite significant progress in automatic SLR in recent years, PSL remains a scarce resource in the field. The development of intelligent systems that support DHH students in specialized science, technology, engineering, and mathematics (STEM) fields. This scarcity limits the development of intelligent systems that support DHH students in STEM education. In particular, the lack of complex PSL datasets and advanced modelling leads to a research gap in capturing the complex hand dynamics associated with complex PSL gestures.

To address this limitation, the study uses a variational autoencoder (VAE) to learn an unsupervised latent representation of hand movement dynamics in PSL. Based on this embedding, a Siamese vision-based transformer (Siamese-ViT) is trained using a triangulator loss function to enhance the model's discriminative ability and improve recognition accuracy. This hybrid approach seeks to overcome data scarcity while ensuring accurate generalization for PSL recognition in real-world educational settings.

1.3 Questions of the Study

- **RQ1:** How well does the hybrid proposed model (Siamese-ViT and β- VAE) based PSL gesture recognition device perform in real classroom conditions for DHH students?
- **RQ2:** How does combining LLM+RAG with the first five alternatives enhance formative learning compared to basic feedback?
- **RQ3:** Is the proposed system stable and appropriate for the Palestinian context?

1.4 Study Aim

This study aims to develop an advanced AI system, Sign Pulse, which uses a hybrid deep learning framework that combines a Siamese variable Autoencoder (VAE) for robust latent representation learning (ViT) for accurate recognition of PSL signs related to STEM concepts.

1.5 Objectives of The Study

The study seeks to achieve the following objectives to reach this goal:

1. Develops a structured dataset for PSL that covers basic mathematical and scientific concepts, with well-annotated videos to support model training and evaluation.
2. Design a hybrid recognition framework by combining a variational autoencoder (VAE) for learning latent representation with a Siamese-Vision Transformer (Siamese-ViT) trained using triplet loss to increase the accuracy and robustness of PSL recognition.
3. Train, fine-tune, and evaluate the proposed recognition model using multiple performance metrics (e.g., accuracy, EER, AUC, and K-accuracy) to ensure effective generalization across PSL gestures.
4. Integrate the recognition model into an interactive educational application, combining PSL recognition with RAG (retrieval- Augmented Generation) and GPT-based models to provide dynamic explanations, personalized feedback, and intelligent tutoring in STEM subjects.

5. Validate the integration of the Sign Pulse system in a real educational environment setting, through expert evaluation and usability testing, ensuring its accessibility, effectiveness, and alignment with the learning needs of DHH students.

1.6 The Significance of the Study

This study gains its significance from its contribution to bridging the educational technological gaps faced by DHH students, particularly users of PSL. Traditional educational systems rely primarily on textual and audio content. This limits these students' access to educational information, particularly in scientific and technical subjects that require visual and interactive explanations. From this perspective, the student believes that developing an advanced AI system capable of automatically recognizing PSL signs and integrating the system into an interactive educational platform provides an equal educational opportunity for this group.

The innovative approach offers an interactive teaching method that enables DHH students to learn more effectively through visual learning, providing immediate responses that contribute to improved comprehension and understanding. The integration of AI and interactive learning technologies represents a significant leap in comprehensive educational tools that cater to the linguistic and cultural needs of deaf and hard-of-hearing students (Trajkovski & Hayes, 2025). By creating a specialized dataset for PSL, a significant step can be taken toward documenting this language and promoting its use in various future applications.

Enhancing Accessibility to STEM Education. The lack of educational resources for deaf and hard-of-hearing students in STEM education is a significant challenge hindering their learning. These disciplines rely heavily on abstract symbols and complex concepts. Communication information becomes more difficult without the use of visual aids or sign language translation. To bridge the gap, the study aims to improve access for deaf and hard-of-hearing students to learn science, technology, engineering, and mathematics (Bonvillian et al., 2020).

Advancing AI Research in Low-Resource Sign Language. The study aims to enhance the integration of PSL into AI systems by developing deep learning-based

solutions that improve the quality of education for the hearing impaired. To achieve this, the study focuses on constructing a specialized dataset that encompasses educational gestures, providing a robust digital foundation for evaluating AI models. Additionally, it seeks to fine-tune the model to accurately classify PSL gestures, enabling the development of intelligent educational tools that assist deaf and hard-of-hearing students in comprehending scientific and mathematical concepts. Furthermore, the study explores the potential of deep learning in analyzing sign language as a non-verbal communication system, paving the way for AI-driven applications that can interpret and respond to PSL gestures in real-world educational settings. By pursuing these objectives, the study contributes to advancing research in low-resource sign languages and enhances opportunities for the adoption of AI technologies to facilitate effective communication for individuals with hearing impairments, achieving greater integration and inclusivity in education.

Supporting Human-Computer Interaction HCI for Deaf and Hard-of-Hearing Students. This study extends beyond static PSL recognition by developing an AI-powered interactive application to enhance human-computer interaction for deaf and hard-of-hearing students. The application relies on real-time PSL sign recognition. Students receive feedback on their academic performance, and the system includes interactive generation by LLMs, helping deaf and hard-of-hearing students participate effectively in learning. This provides seamless interaction between students and smart devices and creates a more dynamic learning experience, making self-learning easier, more convenient, and more efficient in response to individual needs, thus contributing to bridging the communication gap and enhancing educational opportunities for individuals with disabilities, specifically the deaf and hard of hearing.

Potential for Real-World Implementation. The study provides a practical model for enhancing interactive learning for deaf and hard-of-hearing students, with significant implications for educational institutions and assistive technology developers, by developing an AI-based tool that can be integrated into classrooms and special education programs to recognize PSL gestures, this provides a more efficient learning environment and contributes to improving students- teacher interaction within classrooms, specifically in mathematics and science concepts. In addition to the educational benefits of the study, it lays a broad foundation for AI applications, as the

scope of PSL recognition can be expanded to include healthcare and communication across different environments, thereby enhancing opportunities for social integration for deaf and hard-of-hearing individuals.

Bridging the Digital Gap in Deaf and Hard-of-Hearing Education. The absence of AI-powered assistive tools specifically designed for deaf and hard-of-hearing students in PSL exacerbates the learning gap, limiting their ability to engage effectively with digital learning environments. The study aims to address this challenge by developing an AI system for recognizing PSL gestures, ensuring that students with hearing disabilities have equal educational opportunities. By integrating smart technologies into special education, the study enhances digital accessibility, enabling students to interact with educational materials in ways that align with their linguistic communication needs. Furthermore, the study emphasizes the importance of innovation in the field of AI-inclusive education, highlighting the need for continued research and investment in the development of assistive learning technologies. By demonstrating the potential of AI to enhance inclusive education, this study encourages educators, policymakers, and technology developers to prioritize the development of smart, adaptive educational tools that respond to the needs of diverse learners. Through these efforts, the study not only contributes to promoting digital inclusion but also advances a more equitable, technology-driven educational approach to supporting deaf and hard-of-hearing students.

1.7 Limitations of the Study

Despite its contributions to the development of AI-based educational tools to support DHH students, the study faces some limitations, including:

1. The study focused only on signs specific to science and mathematics subjects, which means that the results may not apply to PSL signs used in everyday conversation or other domains, such as social communication or specialized terminology outside of STEM.
2. Data size and diversity: Although the created dataset is comprehensive for educational signs, it is smaller than the large multilingual dataset used for sign languages such as ASL and BSL. This may affect the adaptation to new or rare signs not covered during training.

3. Real-time camera gesture recognition via camera input: The system relies on camera-based input for recognizing PSL gestures in real time. This introduces challenges related to camera quality. This can lead to reduced gesture recognition accuracy in low-light environments or when visual obstructions affect gesture clarity.
4. Limited availability of PSL video-based datasets: PSL suffers from a lack of large-scale, publicly available datasets in video format, which may affect the model's ability to generalize and recognize gestures that are not present in the training dataset. Although a custom dataset was created for this study, the limited diversity of recordings and gestures may impact the model's performance when encountering new signs or dialectal variations within PSL.
5. Lack of extensive data to training data for all usage scenarios: Some challenges remain in covering all possible usage scenarios, such as differences in sign performance speed between users, subtle changes in sign performance, and the impact of environmental factors on the quality of recordings. Therefore, the system may require further modification and continuous improvement to ensure accurate sign recognition in various conditions.

1.8 Structure of Study

This study includes : (a) an introduction outlining the research background, objectives, and significance, (b) a literature review exploring previous research on sign language recognition, AI in education, and inclusive learning for DHH students. (c) Methodology detailing (Model development and dataset construction, focusing on building a structured dataset for PSL, and designing a hybrid deep learning framework based on a variational Auto Encoder, (c) the findings, discussion, and conclusion, which present the experimental results, interpret the model's performance, and offer recommendations for future work and educational applications.

1.9 Hypotheses

Based on the research objectives and conceptual framework of this study, a set of hypotheses was developed to examine the relationship between the proposed variables. These hypotheses aim to verify the effectiveness of the developed AI-based sign language recognition system (Sign Pulse) in improving accessibility, educational

performance, and interaction for DHH students in the STEM education context. Three main hypotheses have been established and will be analysed and discussed in detail in the discussion and recommendations section to determine the educational impact and practical implications of the proposed system.

1. **Null hypothesis (H0):** The hybrid proposed model does not achieve statistically significant accuracy above the deployment threshold (ACC<90%, AUC<0.90, EER >8%). **Alternative hypothesis (H1):** The proposed hybrid model does achieve statistically significant accuracy above the deployment threshold (ACC≥ 90%, AUC≥ 0.90, EER ≤ 8%).
2. **Null hypothesis (H0):** The LLM+RAG with To-5 alternatives does not improve formative learning compared to the baseline system. **Alternative hypothesis (H1):** The LLM+RAG with To-5 alternatives does improve formative learning compared to the baseline system.
3. **Null hypothesis (H0):** The system is not suitable for Palestinian classrooms and fails to achieve acceptable contextual alignment coverage of Palestinian vocabulary. **Alternative hypothesis (H1):** The system is suitable for Palestinian classrooms and achieves an acceptable level of contextual alignment coverage of Palestinian vocabulary.

1.10 Operational Definitions

- **Palestinian Sign Language (PSL):** is a visual means of communication method developed and used by deaf and hard-of-hearing people in Palestine, which consists of hand gestures, facial expressions, and body movements to convey meaning (Al-Fityani,& Padden., 2010). PSL refers to a set of mathematical and scientific signs, recorded and labeled in a custom dataset for training.

- **Deaf People:** deaf people have a hearing loss of 70 decibels or more, making it difficult for them to understand speech through hearing alone, with or without hearing aids (2022 ,عبد العزيز الخضير). In this study, “Deaf” refers to school children with severe hearing loss of 70 decibels or more, which limits their ability to understand speech through hearing alone, even when using hearing aids. The research focuses on students who rely on PSL as their primary means of communication within the educational environment.

- **Hard of Hearing:** these individuals have a hearing loss ranging from 35 to 69 decibels, resulting in difficulty understanding speech through one ear, with or without hearing aids (2022 ,عبد العزيز الخضير). In this study, "hard of hearing" refers to students with hearing loss ranging from 35 to 69 decibels, which impacts their ability to understand verbal educational content in traditional classrooms. These students are identified based on formal hearing assessments and can benefit from hearing aids and assistive technology to enhance their comprehension of educational materials.
- **Retrieval Augmented Generation (RAG)**: is a hybrid framework that combines a retrieval system with a generative language model, allowing the model to access and leverage external knowledge bases to produce more accurate and contextually relevant responses (Lewis et al., 2020). RAG was implemented to enrich system responses by retrieving STEM-relevant explanations and definitions associated with recognized PSL gestures before generating natural language feedback.
- **A Variational Autoencoder (VAE)** is a deep learning generation model that encodes input data into a probabilistic latent space and reconstructs it to capture the underlying data distribution, often used for unsupervised learning and feature generalization (Kingma et al., 2014). The VAE is applied to learn representations of PSL gesture features, improving the generalization of the model and enhancing classification performance across various gestures.

- **Deep learning (DL):** a subset of AI that employs artificial neural networks with multiple layers to learn from large amounts of data, improving classification and recognition capabilities (LeCun et al., 2015). Operational Definition: This study applies deep learning techniques to process PSL gesture images, enabling the model to classify mathematical and scientific signs with high accuracy.

- **Interactive Learning:** a pedagogical approach that fosters active student engagement through technology-enhanced, interactive educational experiences(Antia, 2013) . Operational Definition: The AI system in this study provides real-time feedback by integrating PSL recognition with ChatGPT, allowing students to interact with STEM educational content dynamically.

- **Dataset:** A structured collection of labeled data used for training and evaluating machine learning models, essential for improving AI model performance (Goodfellow et al., 2016) .The dataset in this study consists of video frames of PSL gestures, specifically curated for mathematical and science concepts, annotated and processed to enhance AI recognition accuracy.

- **Large Language Models LLMs**: advanced AI-driven linguistic models trained on vast text datasets to generate coherent and contextually relevant responses(Keloharju & Keloharju,2025), (Brown et al., 2020) integrate LLMs like ChatGPT to provide context-aware explanations and adaptive responses after PSL gestures are recognized, enhancing educational engagement.

- **STEM (Science, Technology, Engineering, and Mathematics)** is an integrated educational framework that combines science, technology, engineering, and mathematics to develop learners' design thinking(Xie et al., 2015) . in this study, STEM refers to the educational content related to mathematics and science that is taught using PSL. The AI model is trained specifically to recognize PSL gestures corresponding to STEM concepts, facilitating interactive learning for deaf and hard-of-hearing students.

- **The Siamese Vision Transformer (Siamese ViT)** is a dual-branch architecture that combines transformer-based visual feature extraction with a siamese contrastive learning mechanism to measure similarity between image pairs (Chen et al., 2021) In this study, Siamese ViT is used to compare PSL gesture embeddings in pairs, enabling accurate recognition and verification of similar or identical signs, enhancing recognition consistency and robustness.

Chapter Two: Literature Review and Theoretical Background

This chapter provides a systematic and critical review of previous studies on the application of AI to support the DHH populations, focusing on ArSL modelling techniques. Despite significant progress in ArSL recognition, a notable absence remains in peer-reviewed research focusing on PSL. This gap highlights the novelty and importance of the study, which seeks to fill this research gap by developing a specialized framework for recognizing PSL and a new dataset.

2.1 Sign Language in the Educational Context

Previous studies have demonstrated that sign language is an effective means of enabling deaf students to engage with academic content emotionally. Sign language is not just a means of communication; it is a cultural and cognitive framework through which many deaf and hard-of-hearing people experience the world. Its role in education has been widely studied, particularly in learning outcomes in STEM education. Deaf students do not learn in a fundamentally different way from their hearing peers, but they need educational support that is linguistically and visually appropriate (Marschark, 2012). They analyzed the academic performance of deaf and hard-of-hearing students. They found that classroom accessibility, including sign language interpretation and visual materials, was a key factor in their success (Antia, 2009). Luckner et al. (2001) emphasized using visual aids and sign-based explanations when teaching abstract STEM content. Adopting a bilingual education model, which combines sign language and written language, enhances the academic achievement and social integration of deaf students in Arab countries. This model highlights the importance of providing an educational environment that supports both languages to achieve better educational outcomes (Kubicek, 2021). There is a strong correlation between sign language skills and writing skills among deaf students. Enhancing proficiency in sign language can positively support and enhance literacy skills, supporting the effectiveness of bilingual education models in improving the academic achievement of these students (Zhang et al., 2024). Shifting from the spoken language-based teaching approach to bilingual and bicultural models, their findings are consistent with the studies conducted in Finland and Japan, where translating scientific textbooks into sign language significantly improved comprehension and academic engagement (Knoors, 2014). The most

common strategy teachers use with students with hearing impairments is total communication, which combines sign language, oral reading, and written communication by attracting the student's attention before commencing communication to ensure effective interaction (Sultana et al., 2023). The integration of sign language into the educational process demonstrates that it is not merely a means of communication but an essential element for achieving linguistic equality and educational inclusion for DHH students.

2.2 Automatic Sign Language Recognition

Automatic sign language recognition SLR involves automatically identifying signs from video input, typically using computer vision and deep learning. This is a crucial step toward enabling sign language-based interactive systems for educational and accessibility purposes. Figure 1.2 presents the general workflow of automatic SLR systems, providing a visual overview of the main stages in building the system.

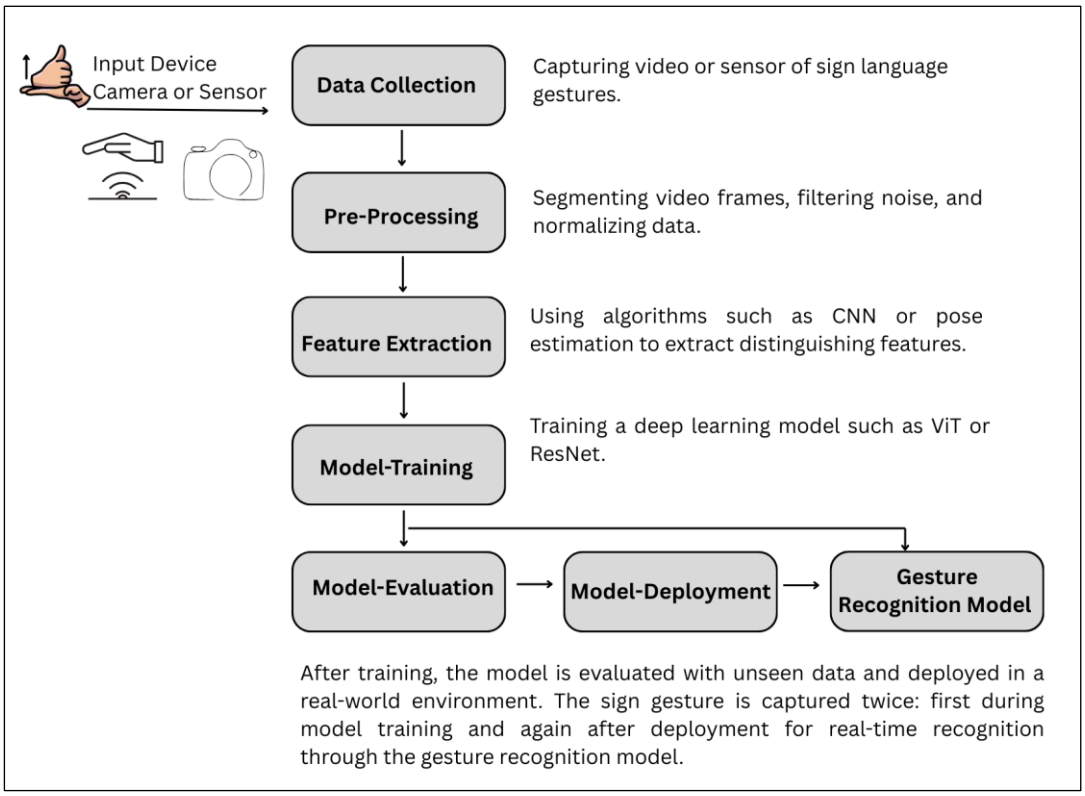


Figure 1.2: General Workflow of an Automatic Sign Language Recognition System.

Camgoz et al. (2018) proposed a comprehensive neural architecture using sequential learning with attention, enabling translation from gesture videos into written sentences. Studies the real-time requirements of the SLR systems using artificial neural networks, ANNs, and hidden Markov models (HMMs), focusing on the Brazilian sign language LIBRAS, and focusing on the balance between system speed and accuracy (Cooper, 2013). Singha (2013) addressed Indian Sign Language (ISL) recognition in live video by detecting skin regions, applying histogram matching, and utilizing eigenvalue-based feature extraction. Classification relied on a weighted Euclidean distance technique. Mohandes (2014) addressed the lack of research on Arabic sign language, ArSL, using Gaussian skin tone modelling and hidden Markov models (HMM), and developed the first Arabic sign language recognition systems for isolated signs, proving a foundation for Arabic-speaking communities. Pigou et al. (2015) proposed a deep convolutional neural network model for fingerspelling recognition in American Sign Language ASL, and this approach achieved real-time classification without the need for manual feature extraction, representing a shift towards fully automated SLR. Rastgoo (2021) introduced Multiview skeleton-based recognition using deep learning. Their dataset featured multiple camera angles to enhance recognition across varied visual conditions.

Camgoz et al. (2020) proposed a transformer-based system for continuous German sign language DGS recognition and translation, representing a shift towards sequential modelling with deep learning and contextual learning. Singh (2022) reviewed the intelligent SLR methods for Indian Sign Language ISL, analyzing rule-based, sensor-based, and learning-based systems, highlighting gaps and future directions. In this study, the researchers conducted a comprehensive study focusing on recent technological developments aimed at improving communication for individuals with hearing and speech impairments. The researchers reviewed innovations such as smart gloves, Android applications, and technologies such as convolutional neural networks CNNs, Gaussian filtering, and models (HMMs), and speech-to-text techniques (Naresh et al, 2020). (Madhiarasan, 2022) Conducted a systematic review of sign language recognition SLR systems, categorizing the current literature based on input methods, dataset types, algorithms, traditional machine learning, deep learning models, and key performance. Pu et al. (2020) Their method improved model alignment and recognition accuracy by simulating substation insertion, detection operations, and video-text pairs to bridge the gap between the CTC training loss and the true performance metric, the

WER word error rate. Noor et al. (2024) developed a hybrid deep learning model combining CNN and LSTM to recognize Arabic sign language gestures. Two recent studies have made significant contributions to the development of Arabic sign language, using the RGB-based AASL dataset, and their model achieves high accuracy in ArSL recognition, confirming the effectiveness of deep CNNs in classifying isolated Arabic letters (El Kharoua, 2024). Abu-Jamie et al. (2022) explored and compared the performance of Mobile Net and VGG16 architectures on a local ArSL dataset.

Table 1.2: Summary table of reviewed sign language system studies

Study	Sign language	Input type	Methodology	Key contribution
Camgoz et al. (2018)	RWTH-PHOENIX-Weather	Video (RWTH-PHOENIX-Weather)	End-to-end Neural Machine Translation (NMT) for CSL	Introducing the first NMT-based SL translation system
Cooper (2013)	British Sign Language (BSL)	Video (3D Hand Trajectory)	CNN for 2D/3D hand trajectory	Emphasized 3D hand pose in continuous SLR
Singha (2013)	Indian Sign Language (ISL)	Skeleton data from video	DL on hand skeletons	Multiview improvement for gesture recognition
Mohandes (2014)	Arabic Sign Language (ArSL)	Hand Shape Image	SVM on extracted hand shape features	Classical ML for isolated SLR recognition
Pigou et al. (2015)	RWTH-PHOENIX	Video frames (RWTH-PHOENIX)	CNN + LSTM	Real-time SLR using CNNs on video frames
Rastgoo (2021)	Multiple languages, American Sign Language (ASL), and Chinese Sign Language (CLS)	RGB and Depth video	CNN-RNN with transfer learning	Cross-lingual SL model using RGB and depth
Camgoz et al. (2020)	PHOENIX14T	Video (PHOENIX14T)	Transformer + BERT	Continuous SLR and sentence-level translation

Singh (2022)	ISL	Video frames	3D-CNN and LSTM	Indian ISL with spatial-temporal features
Naresh et al. (2020)	No specialized dataset (general SLR)	General video data	Hybrid deep learning (CNN-RNN)	Improved ISL gesture classification
Madhiarasan (2022)	No specialized dataset (general SLR)	General video data	Residual CNN with attention	Improved real-time ISL recognition
Pu et al. (2020)	CSL	Skeleton sequence (graph-based)	ST-GCN (Spatial Temporal Graph ConvNet)	Applied GCNs to sign gesture dynamics
Noor et al. (2024)	ArSL	Image and Video	Hybrid CNN-LSTM	Hybrid CNN-LSTM model for Arabic SLR.
El Kharoua (2024)	ArSL	RGB Images	CNN-based model for recognizing the ArSL alphabet signs	using the RGB-based AASL dataset
Abu-Jamie et al. (2022)	ArSL	Images	Mobile Net and VGG16 for ArSL	explored and compared the performance of Mobile Net and VGG16 architectures

The studies in **Error! Reference source not found.**2 The above shows a clear evaluation in sign language recognition SLR research, from traditional machine learning approaches, e.g., SVM in Mohandes (2014) to deeper learning, such as CNN-LSTM hybrids (Pigou et al. 2015) and transformer-based models (Camgoz et al., 2020) Each study contributed uniquely to the advanced technical capabilities of the SLR system, such as improving gesture recognition accuracy (Singha 2013), and enabling real-time detection of sign language (Pigou et al., 2015) However, the key observation in these studies is the limited focus on educational environments. While the methodologies are promising in controlled settings, most systems have not been expanded or adapted to support learning contexts, particularly for deaf and hard-of-hearing students. The lack of attention to curriculum integration, learner interaction, and real-time tutoring reflects a wide gap in applying SLR research to an inclusive

pedagogical system. None of the mentioned works address low-resource sign languages, such as PSL, nor present frameworks specially designed for STEM education. The comparative analysis highlights the originality of this thesis. Unlike previous studies, this research bridges the gap between sign language recognition and intelligent educational support, creating a classroom-ready system that combines ViT-based self-learning with LLM-supported tutoring.

2.3 Building a Specialized Sign Language Dataset

One of the most significant constraints facing the development of accurate sign language recognition systems is the availability of high-quality datasets. Despite vast textual corpora, sign language resources remain significantly limited. Particularly those representing non-Western sign languages or educational vocabulary. Developing effective sign language recognition SLR systems requires the construction of high-quality, diverse datasets that reflect actual linguistic usage. Recent studies have significantly addressed this need across various sign languages. (Othman, 2024) proposed a structured framework for creating sign language datasets, emphasizing a comprehensive process that begins with participant recruitment and obtaining ethical approval, continues with multimedia video recording, and concludes with accurate annotation using tools such as ELAN. Provided a valuable resource by compiling the sign language dataset, which unifies information across over 40 existing datasets. Detailing aspects such as the sign language used, the number of samples. Illustration methods and media types (Kopf et al., 2022). Presented the SCOPE dataset for Chinese Sign Language CSL, which contains over 72 hours of contextual conversational videos and nearly 60,000 captions (Liu et al. 2025). Alishzade (2025) developed the ArSLD dataset for Azerbaijani sign language, including fingerspelling recordings, isolated words, and complete sentences, and the dataset includes over 30,000 video samples, providing the first open source for Azerbaijani sign language and laying the foundation for resource-limited languages.

Jiang (2022) introduced SDW-ASL, a large-scale, continuous American Sign Language ASL dataset containing over 416,000 words and 30,000 sentences, generated through an automated system. Rodriguez et al. (2025) presented a lightweight and practical dataset for Mexican Sign Language LSM, covering 29 letters and 10 digits, using tools such as Media Pipe for feature extraction and evaluation of multiple

classifiers, KNN, CNN, and RNN. The dataset supports experiments on small-scale SLR systems. To illustrate the diversity of methodologies and the contribution of specialized sign language datasets, Table 2.2 provides a comparative summary of key studies. The table highlights important aspects, such as the target language, the size and duration of the dataset, the tools and techniques used, and the unique significance of each study.

Table 2.2: Comparative summary of key studies

Study	Language	Size/Duration	Key methods	Contribution/Significance
(Othman, 2024)	General framework (all languages)	Not specified (theoretical framework).	ELAN, participant recruitment, ethical design	Provides comprehensive guidance for building SL datasets.
Kopf (et al., 2022).)	Multiple (meta-compendium)	40.000 datasets.	Standard metadata for existing senses	Enables comparison and selection of existing SL resources.
Liu et al. (2025)	Chinese Sign Language CSL	72 hours, 59,231 captions	Contextual video Conversational scenes	Enhances contextual modelling in CSL.
Alishzade (2025)	Azerbaijani sign language, ArSLD	30,000 video samples.	Manual annotation, two camera angles.	First large-scale open-access ArSLD resource.
Jiang (2022)	American Sign Language ASL	104 hours 416000 words 30,000 sentences.	Automated system, synthetic generation.	Demonstrate the scalability of a machine-generated dataset.
Rodriguez et al. (2025)	Mexican Sign Language LSM	29 letters 10 digits	Media pipe, KNN, CNN, RNN evaluation	Supports small-scale experimental SL systems

After reviewing the summarized studies in Table 2.2 and earlier, several thematic insights emerge that highlight the progress and limitations in the field of sign language dataset construction. Taken together, these studies represent a fundamental

effort to diversify low resources across different sign languages, including low-resource contexts, such as Azerbaijani Sign Language (Alishzade, 2025), and Mexican Sign Language (Rodriguez et al., 2025) They also reflect increasing methodological sophistication, such as the use of automated data generation (Jiang,2022) and contextual video annotation (Liu et al., 2025). However, there are important gaps that directly enrich this thesis. While previous work provides valuable data resources, most still focus on gesture recognition in isolation, with limited integration of full-sentence-level communication or instruction. It is worth noting that few studies address the contextual richness required for real-world learning environments such as STEM education in deaf and hard-of-hearing classrooms. Kopf et al., (2022).

Despite presenting a comprehensive set of datasets, the field still lacks uniform standards for collecting, classifying, and sharing data. This lack of consistency makes it challenging to develop AI models that can be easily modified or reused across various sign languages or educational applications. None of the previous datasets includes Palestinian sign language, reinforcing its status as a low-resource language system in AI research. The absence of Palestinian sign language limits regional inclusion and contributes to systematic disparities in educational technologies available to Palestinian deaf learners. Furthermore, (Jiang,2022) and (Liu et al., 2025) have only begun to move towards dynamic and conversational datasets, but these are still limited to American Sign Language and Chinese Sign Language, respectively. It is not directly related to curriculum content or classroom application. Therefore, a significant research gap lies in developing dominant-compatible, culture-informed sign language datasets that can serve as the backbone of AI-powered educational platforms.

2.4 Artificial Intelligence Application in Education

In recent years, the integration of AI into the educational landscape has brought about transformative changes in enhancing efficiency and enabling more intelligent, personalized learning environments. Among the most impactful innovations are large language models LLMs such as ChatGPT, which have demonstrated remarkable potential in supporting adaptive learning, producing personalized educational content, and facilitating natural language-based interaction between students and systems. These developments are particularly important when applied to support learners with spatial needs, such as deaf and hard-of-hearing students.

Artificial intelligence opens new horizons for enabling sign language-based communication, multimodal interaction, and more inclusive instructional design. **Error! Reference source not found.** illustrates a general framework for building sign language systems for AI-powered educational applications. It illustrates the basic steps, from input method to feature extraction and learning architecture, and finally to sign language representation or translation. This section reviews recent studies exploring the role of AI in educational settings, with a focus on systems designed to enhance the learning experiences of students with hearing impairments.

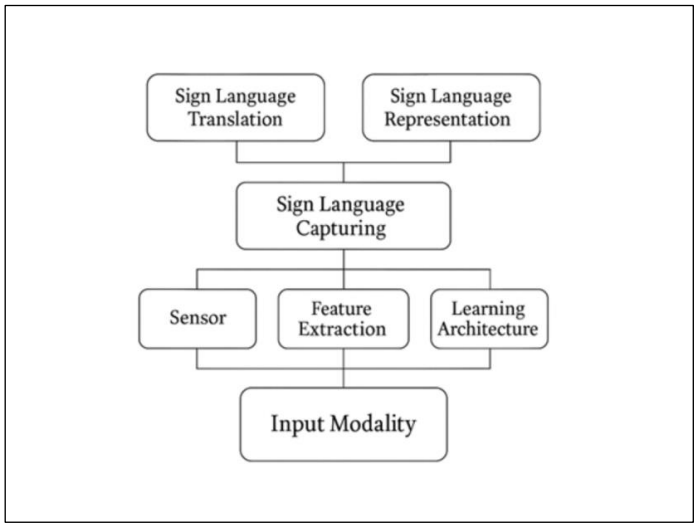


Figure 2.2: A framework for developing sign language technologies used in educational AI applications

(Cheng, 2024) Conducted pioneering how deaf and hard-of-hearing learners interacted with AI tutors powered by large-scale language models and designed with different teaching personas, their results showed that learners were more engaged and confident when AI tutors reflected on their experience teaching deaf and hard-of-hearing learners, highlighting the importance of cultural relevance and transparency in AI-based instructional design. Integrating the Internet of Things (IoT) with augmented reality AR to create an AI-powered learning platform specifically designed for deaf users is cutting-edge technology in education in general and special needs education in particular (Alrashidi, 2023).

(Arroyo Chavez, 2024) Demonstrated the use of large language models LLMs in personalized interpretation, their system allowed for more accurate and contextually

appropriate translations, directly addressing common barriers faced by deaf and hard-of-hearing students when interacting with audiovisual educational content. Providing an interdisciplinary perspective on sign language recognition, generation, and translation, focusing on the technical and educational impact of AI systems in supporting communication accessibility, the study discussed how these systems can integrate into learning environments to empower learners with disabilities (Bragg et al. 2019). Stinson et al. (2022) explored the impact of training in messaging and communication strategies on interactions between deaf and hard-of-hearing college students in group work settings, in designated educational tools that support effective interaction and equitable communication for learners with disabilities. Evaluated the use of automated speech recognition ASR technologies to generate real-time translations in higher education classrooms, but also emphasized the option of AI systems to improve access to lectures in higher education (Butler et al. 2019). Camgoz et al. (2018) presented one of the earliest end-to-end neural machine translation models for sign language. The system was trained on continuous video data of sign language users performing German Sign Language (DGS), and the inputs were translated into spoken sentences. By leveraging deep learning architectures, particularly sequence-to-sequence models with attention mechanisms, the study demonstrated the feasibility of generating full-sentence-level translations directly from sign language videos. This work laid an important foundation for subsequent research in sign language translation, although it didn't focus on educational applications or interactive feedback. To better understand the reality of AI applications in educational contexts for the deaf and hard of hearing, Table 3.2 provides a comparative analysis of the most significant studies. Each study is evaluated based on the nature of the developed application, the technologies used, the observed educational impact, and the identified research gap. This structured overview allows us to track the successes of current efforts and the remaining shortcomings.

Table 3.2:Summary of studies for developing sign language technologies

Study	Application Developed	Impact/Findings	Modality	Gap Identified
Cheng (2024)	LLM-powered AI tutoring	Tutors with experience working with deaf and hard-of-hearing	LLMs, user personas,	Limited to perceptual

	system for the deaf and hard-of-hearing-aware personas	individuals were perceived as more trustworthy and effective by those who are deaf or hard of hearing.	and educational chatbots.	analysis, lacks integration with sign recognition, or full educational feedback.
Alrashidi (2023)	AI and augmented reality AR platform for the deaf and hard of hearing e-learning	Improved interactive experience with sign language prompts.	IoT+ AR	No real-time tutoring or assessment features.
(Arroyo Chavez, 2024)	LLM-customized closed captions.	Enhanced suitable accuracy and contextual adaptation.	LLM, such as GPT for captions.	Focused on media captions, not full educational interaction.
Bragg et al. (2019)	Framework for SLR recognition generation and translation.	Established an interdisciplinary basis for AI use in sign language.	Mixed AI recognition, generation, and translation.	Lacks integration with feedback or classroom platforms.
Stinson et al. (2022)	Communication training model for mixed deaf and hard-of-hearing.	Boosted team communication and equity for deaf and hard-of-hearing students.	Massaging strategies and qualitative analysis.	Not an AI-based application, but interaction training.
Butler et al. (2019)	ASR system for captioning university lectures.	ASR helps with accessibility despite the accuracy limit.	Automatic speech recognition.	No sign language input or recognition.
Camgoz, et al (2018)	Natural Sign Language translation system.	Enabled sentence-level SLR translation with deep learning.	Natural machine translation NMT.	Doesn’t include personalized educational content.

As shown in the Comparative Table 3.2, existing research demonstrates various attempts to apply AI to benefit the deaf and hard of hearing, starting with an LLM-

powered tutoring system (Cheng, 2024). To augmented reality-enhanced e-learning environments (Alrashidi, 2023). However, most of these systems either focus on specific methods, such as closed captioning or automatic speech recognition, or address isolated challenges, such as communication strategies or translation, without offering a unified learning solution.

2.5 Recent Advances in VAE Architectures

Recent years have seen radical advances in VAE architectures applied to sign language, with researchers moving from traditional models to sophisticated hybrid architectures that integrate attention and reinforcement learning mechanisms. β-VAE models represent one of the most notable developments, offering fine-grained control over feature disentanglement through a β parameter that regulates the balance between reconstruction accuracy and the level of abstraction in latent space (Wang et al., 2024). The researcher suggests that combining transformer structures with VAE achieves superior results in processing long movement sequences, as self-attention mechanisms enable the model to capture long-range dependencies between different parts of the linguistic movement (Havrylovych et al., 2023). This development is particularly important in sign language, where linguistic units can extend across long temporal periods and require a comprehensive understanding of context. The integration of the attention mechanisms into VAE frameworks has proven highly beneficial for sign language applications. Multi-head attention enables the model to focus on different aspects of movements simultaneously, including hand shape, movement trajectory, and facial expressions. This parallel processing capability significantly improves the model's ability to capture the multi-modal nature of sign language communication.

2.6 Advanced Models and Contemporary Applications

Zhou and colleagues introduced the Signs as Symbols (SOKE) model, which represents a paradigm shift in sign language generation using VAE(Zuo et al., 2024). This model utilizes a decoupled tokenizer that converts continuous movements into a token sequence representing distinct body parts, allowing for more precise and flexible processing of motion data. The model also employs a multi-head decoder mechanism, which enables the simultaneous prediction of multiple tokens, thereby improving inference efficiency while maintaining effective information fusion across different

body parts. SOKE architecture introduces several innovative components that address long-standing challenges in sign language processing.

The decoupled tokenization approach enables the independent modelling of different body parts while preserving their coordinated relationships through learned attention mechanisms. This design philosophy enables the model to capture both the hierarchical structure of sign language from finger movements to full-body gestures and the temporal dependencies that characterize linguistic sequences. Furthermore, Sinz and Bejarano developed an RVQ-VAE model specifically for the generative processing of sign language poses, addressing the challenges of abrupt translations and poor smoothness in sign language production(Cruz et al, 2024).

RVQ uses residual vector quantization techniques to improve reconstruction quality and create more natural and soothing intermediate frames. The RVQ approach enables multi-level quantization that preserves fine motion details while achieving efficient compression. The RVQ-VAE methodology represents a significant advance in addressing the temporal coherence problem in sign language generation. Traditional VAE models often struggle to maintain smooth transitions between discrete poses, resulting in unnatural or jerky motions that impair the clarity of the generated signs. The residual quantization approach addressed this problem by learning hierarchical representations that capture motion across multiple temporal scales.

2.7 Future Trends and Implications

Based on a critical understanding of pedagogical challenges and technological advancements, the present study argues that the future of sign language technology in education must go beyond simple classification or translation functions. There is an urgent need for educational systems capable of interacting with deaf and hard-of-hearing learners in real time in a pedagogically meaningful and culturally aware manner. From a theoretical perspective rooted in the philosophy of inclusive education, AI-based solutions must reflect students lived experience and actual needs, not as active participants in constructing their learning environments. Based on the four main topics discussed in the literature review: sign language in educational contexts, automatic sign language recognition SLR, building specialized sign language datasets, and AI applications in educational settings, several future trends and implications emerge:

1. **Advancing pedagogical integration of sign language:** Despite increased recognition of the importance of sign language in educational settings, studies still lack concrete strategies for integrating sign language fluency into STEM curricula. Future research should focus on designing a curriculum with teachers of the deaf and hard of hearing and exploring how AI tutors can support conceptual learning through both visual and textual media.
2. **Enhancing real-time automatic sign language recognition:** Despite improved technology in SLR systems, many models are still limited to isolated signs and lack contextual understanding. Future Directions must include advancing continuous sign language recognition and exploring how ViT and transformer-based models can capture the nuances of gestures, shared expression, and user variation in real-time.
3. **Developing representative and ethically collected sign language datasets:** one of the most challenging is the lack of large, diverse, and ethically sourced sign language datasets, particularly in low-resource regions and dialects. Future research must prioritize comprehensive data collection, community engagement, and open sharing.
4. **Creating intelligence and culturally responsive education systems:** Despite the introduction of AI tutors, most of them cannot dynamically adapt to the needs of deaf and hard-of-hearing learners with special needs or provide specialized educational support.
5. **Bridging gaps across modalities:** An important though still unexplored area is the integration of gestures, speech, text, and visual low resources into a unified multimedia learning platform. Future education systems must allow students to navigate the learning process with flexibility**,** switching between educational media as needed, with consistent content. As shown in **Error! Reference source not found.** The proposed PSL-based framework reflects this shift by combining a real-time sign language recognition engine with a STEM-oriented learning environment. The use of vision transformers and vision language recognition enables more accurate gesture understanding, while the integrated AI infrastructure enables adaptive learning loops. This move toward closed-loop, context-aware systems points to a growing focus on dynamic, personalized learning for deaf and hard-of-hearing students. Particularly in underserved linguistic areas, such as PSL.

In response to these gaps, the proposed system introduces several novel contributions: **First,** it is among the earliest academic efforts to design an educational AI system based on Palestinian Sign Language PSL, a language rarely, if ever, represented in current AI datasets or research. This contribution directly addresses linguistic and geographical gaps in the SLR literature, bringing long-overdue representation to the marginalized community.

Second, the system goes beyond the traditional limitations of isolated gesture recognition or simple translation. It provides a fully integrated, classroom-ready environment specifically designed for STEM education, a significant improvement over previous studies that focused largely on general communication without direct alignment with the curriculum.

Third, the architecture leverages the Siamese-ViT for accurate real-time gesture recognition and large language models LLMs for dynamic, pedagogically relevant dialogue. This dual integration enables the system to act as a contextual tutor, able to understand student input, provide meaningful answers, generate follow-up questions, and explain complex math and science topics through sign language. Unlike previous work, which often treated sign recognition and learning as two separate fields, this thesis presents a framework that combines SLR technology, natural language understanding, and a comprehensive teaching methodology into a sign; therefore, the prototype in the figure 2.3 presented here is not only of direct relevance to Palestinian sign language speaking communities but also serves as a scalable, replicable model for low-resource sign languages.

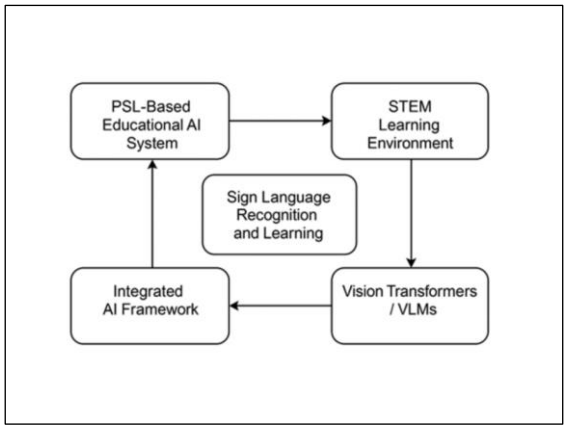


Figure 2.3: Proposed AI-powered educational system for Palestinian Sign Language (PSL), integrating real-time sign language recognition, vision transformers, and a comprehensive STEM learning environment

Chapter Three: Research Design and Methodology

3.1 Research Design

This study details the development of Sign Pulse, an intelligent system for recognizing Palestinian sign language (PSL). The researcher employs an applied development methodology guided by the principles of Design-Based Research (DBR), a suitable approach for developing AI-driven educational technologies within authentic learning contexts. The primary objective of Sign Pulse is to support the teaching of STEM education for deaf and hard-of-hearing (DHH) students in grades one through four, in alignment with the Palestinian curriculum. To practice the DBR farmwork, the researcher implemented an iterative workflow modeled after the input-process-output (IPO) cycle. This model consists of four interconnected stages that form a continuous improvement loop. Importantly, the output from each stage is systematically analysed and fed back into preceding stages. This feedback mechanism is central to the project, enabling the progressive fine-tuning of system standards, the enhancement of the model's recognition accuracy of the model, and the effectiveness of educational interactions by the Sign Pulse application.

The development cycle begins with Data Collection, where authentic PSL classroom videos are recorded. These recordings undergo a preliminary Pre-processing stage, Segmentation, Skeleton Extraction, and Expert Labelling to create clean, machine-readable input. The researcher then uses a Variational Autoencoder (VAE) to learn an unsupervised latent representation of hand movement dynamics. Based on this embedding, a Siamese Vision Transformer (Siamese-ViT) is trained using a Triple Loss. Enabling few-shot recognition of unseen signs. Once the sign is identified, its corresponding interpretation is passed to a Large Language Model (LLM), which generates an educational explanation or question tailored to the student's grade and subject. The content is delivered in real-time, and students' interactions in the classroom (e.g., student responses) are captured as feedback. This feedback is looped back into the pipeline either as mark corrections or as additional samples for the model refinement, thus closing the iterative design cycle. As depicted in **Error! Reference source not found.**, the iterative workflow consists of four interconnected stages that guide the development of the PSL-based educational system: Data Collection, Pre-

processing, VAE Pretraining, Siamese-ViT Few-Shot Learning, LLM Integration, and Classroom Interaction. It highlights feedback paths for label correction and model refinement. This approach combines theoretical foundations with iterative practical implementation.

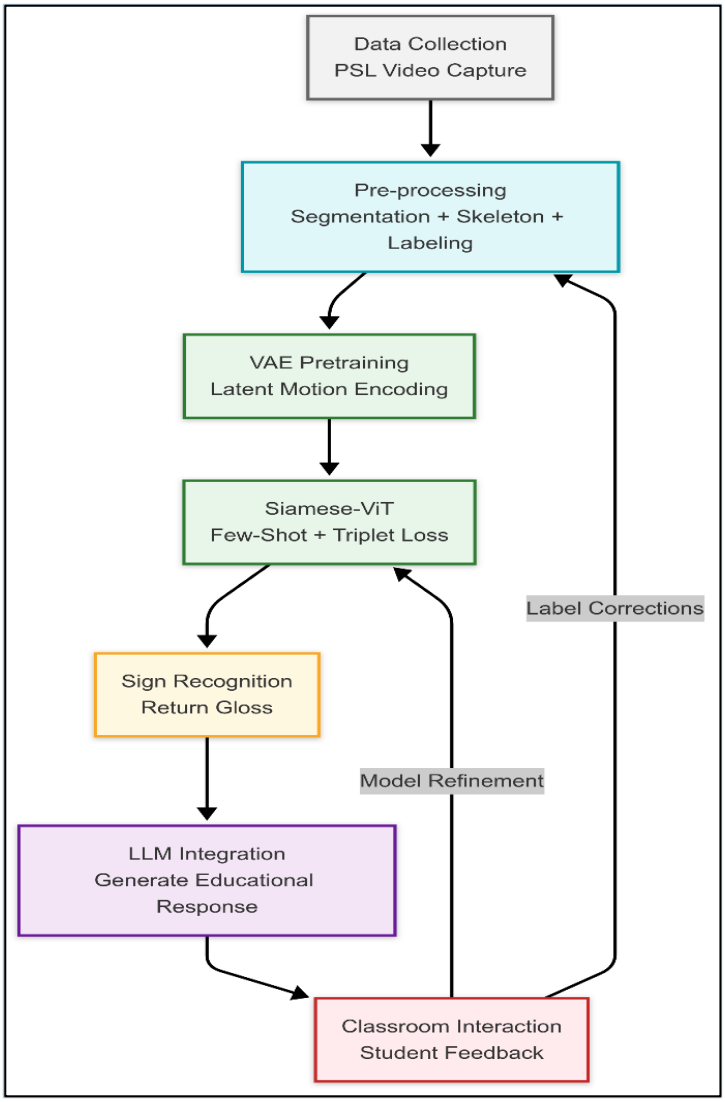


Figure 1.3: workflow diagram illustrating the iterative development cycle

3.2 Study Population and Sample

The study population consists of all DHH attending Palestinian schools in the West Bank and the Gaza Strip across all grades. This population in Palestine represents a significant and underserved segment within the national educational system, facing

various challenges in accessing suitable learning educational content, particularly in academic, mathematical, and scientific subjects, which require abstract explanations delivered through spoken and written language (Khandaqji et al., 2025a).

This population experiences a clear educational gap due to the lack of learning resources adapted for sign language users and the scarcity of interactive, accessible educational content in PSL (Khandaqji et al., 2025b). These challenges necessitate the development of AI-based education systems to bridge this gap and enhance accessibility. Due to the applied nature of the study, the sample didn't focus on human individuals (students) but rather represented a set of PSL for mathematics and science concepts, and general words used in the classroom. These gestures were developed and recorded with the help of native sign language teachers and professional PSL interpreters. The selected concepts are consistent with the official Palestinian curriculum for primary school grades and cover basic areas such as numbers, arithmetic, operations, geometry, shapes, measurements, states of matter, the five senses, basic life science, energy, animal names, general vocabulary, and verbs commonly used in classrooms.

The study seeks to develop a specialized AI application that meets their linguistic and methodological needs. The researcher employs a purposive sampling strategy that ensures both geographic diversity and gender equity. To highlight the magnitude of the problem and justify the focus on students with hearing disabilities, the study relied on recent official statistics. Data from the Palestinian Central Bureau of Statistics (PCBS) indicate that the number of persons with disabilities in Palestine is estimated at approximately 115,000 (2,1% of the population). In 2019-2020, the survey results also show that approximately 12% of children (2-17 years) suffer from at least one functional difficulty. UNICEF reports indicate that children with disabilities, including those with hearing impairments, face severe marginalization and accumulated educational challenges (UNICEF, 2023; ACAPS, 2024; Palestinian Central Bureau of Statistics, 2022). Based on these findings, the percentages for different types of disabilities were compiled and formatted in Figure 3.2 shows the types of disability among children in Palestine.

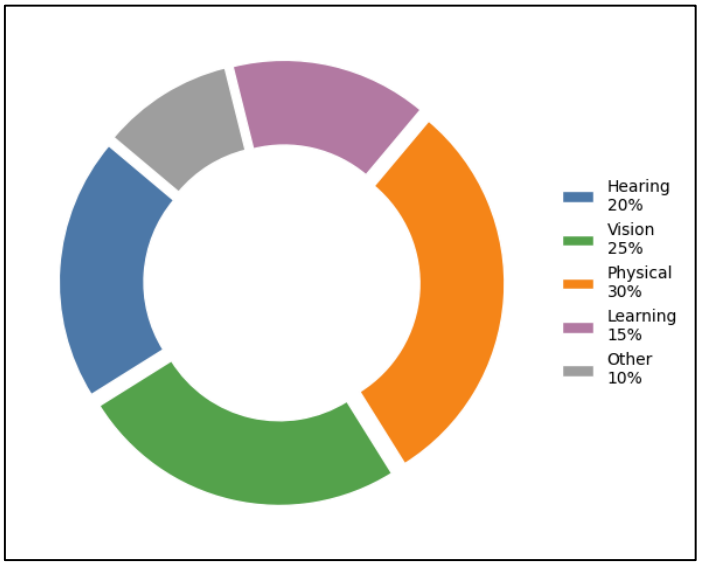


Figure 3.2: 11Proportion of Disability Types among Children in Palestine (Recombined from PCBS, UNCEF, and APCPS Tables.

3.2.1 Target Population

The target group includes all DHH students enrolled in Palestinian schools in the West Bank and Gaza Strip, at all educational levels (grades 1-12), including comprehensive government schools and specialized institutions affiliated with the Ministry of Education and Higher Education. The basic distribution is based on the ministry's administrative statistics (2023/2024). According to the national population, DHH students in 2025 are estimated at approximately 2741 students, of whom approximately 963 are deaf, and the remainder are hard of hearing (partial hearing loss). At this stage of the Sign Pulse assessment, statistical inference is limited to grades one through four, and the sample frame for these grades in 2025 is approximately 963 students.

3.2.2 Sample Frames

The sample frame, restricted in this phase to grades one through four, is constructed in three linked steps:

1. Administrative rosters: compile lists of Ministry of Education for 2023/2024 directories/school listing for DHH students in primary grades one through four

across both inclusive public schools and specialized institutions (State of Palestine, Ministry of Education, 2023).

2. School verification: the content selected school administration to confirm the number of current eligible one through four grades students at DH, obtain contact information for families (for approval and follow-up), and verify classroom and technology readiness for Sign Pulse sessions.
3. One site record review: cross-check advisor, resources room records to apply inclusion, exclusion criteria, remove duplicate entries, across sources, and finalize the eligible list.

Multi-source frameworks provide an up-to-date and reliable basis for relative stratification in the West Bank and the Gaza Strip, with one through four, and reduce the likelihood of missing or duplicated cases. Where appropriate, within school selection, the school may use systematic random or purposive procedures to meet inclusion criteria (State of Palestine, Ministry of Education, 2023).

3.2.3 Sample Design

The researcher adopts a purposive mini-stratified two-stage design to maximize diversity and preserve representatives within grades one through four:

Stage One: School selection. Schools are divided into four cells based on geographic location (North vs South) and social context (Urban vs Rural). One school is selected from each cell based on confirmed enrolment of DHH students in grades one through four, administrative approval, and read lines for Sign Pulse sessions, such as space, technology, and coordination.

Stage Two: in the school selection (partially balanced distribution), the goal of each selected school is to reach approximately 10 students, DHH pupils with an appropriate 50:50 gender balance of about 5 boys and 5 girls, distributed across grades one through four, with an enrolment of two to three pupils per grade to ensure curriculum coverage. The study implements a pre-specified within-school replacement protocol confined to the same stratum (school, gender, grade) using a prepared reserve list, whereby any student unable to participate is replaced by a matching student from the same stratum without changing the sample size or imbalance (Lynn et al., 2004;Demarest et al., 2022).

3.2.4 Sample Size Determination and Statistics Power Analysis

Determining the appropriate sample size is a very important methodological factor in quantitative research design (Cohen, 1988). For the current phase, one through four grades, the working population is estimated at 963 pupils nationwide (Ministry of Education, 2023).

Phase One: Estimating the initial sample size (infinite population Cochran model), using the Cochran formula for proportions, Cochran at 95% confidence. p=0.50 and margin of error e=0.10, (Cochran., 1977):

$$n_0 = \frac{z^2 p(1-p)}{e^2} = \frac{1.96^2 \times 0.25}{0.10^2} \approx 96 \quad (3.2.1)$$

Phase Two: finite population correction (FPC). Adjusting for N ≈ 963.

$$n = \frac{n_0}{1+\frac{n_0-1}{N}} = \frac{96}{1+\frac{95}{963}} \approx 87 \quad (3.2.2)$$

Phase Three: Allowance nonresponse/ attrition allowance. Applying a pragmatic 10 to 15% allowance brings the final target back to n≈ 96, while maintaining the design precision, while accounting for the expected field loss. The selected sample size (n=96) provides sufficient statistical power (1-β ≥ 0.80) to detect medium effect sizes (Cohen's d ≥ 0.50) at α = 0.05, consistent with Cohen's (1988) criteria for behavioural research. The following Table 1.3 illustrates the Population Sample and Sample Size Calculations, and Table 2.3 shows the Mini Stratified Sample Design based on the Ministry of Education 2023 data sheets, with clear statistical and methodological criteria.

Table 1.3. Population Sample and Sample Size Calculations

Item	Value
Study population (current evaluation phase)	All DHH students in grades one through four in Palestine schools (Westbank Gaza).
Baseline/ reference year	Ministry rosters 2023 with a + 7% operational update for 2025.
Sample frame (N)	≈ 963 students
Confidence level (z)	95% (1.96)

Assumed variance (p)	0,50
Margin of error (e)	± 10%
Intimal size (Cochran, n_0)	≈ 96
Finite population correction (FPC)	≈ 87
Expected attrition (10-15%)	Planned minimum target ≈ 96
Implemented the sample in phase 1	≈ 40 pupils.

Table 2.3: Mini Stratified Sample Design

Item	Value
Number of selected schools	Four schools via a four-cell stratification: North-Urban, North-Rural, South-Urban, South-Rural
Pupils per school	10 (total ≈ 40)
Gender balance	Five males and 5 females per school (50:50)
Grade distribution within the school	Two to three pupils per grade to ensure curriculum coverage.
HoH / deaf representation	Preserving the school's actual deaf mix where feasible
Replacement protocol	Within stratum replacement (same school, grade, gender).

3.2.5 Data Collection and Annotation

Data collection represents the foundational steps of the study, as it ensures the model is trained on high-quality, representative samples of PSL gestures specific to the STEM domain. The following section outlines each stage of data preparation. From raw video capture to data formatting for model training. Each process was carefully selected to ensure clear gestures, ensure consistency, and enhance model robustness. Video data were collected under controlled environmental conditions to ensure consistency across samples. Certified PSL interpretation was videotaped during the performance of 434 sign-specific movements. The videos were stored at high resolution to enable accurate feature extraction and subsequent normalization. Table 3.3 below summarizes the tools

and components used to collect and annotate the dataset. The figure of the sample of images from the video sign.

Table 3.3: Collecting PSL STEM Videos

Tools/ components	Descriptions	Justification
PSL-STEM dataset	A 434-video dataset of PSL gestures representing STEM concepts.	Provides curriculum-aligned, sign-specific data for training the recognition model.
Expert's sign language	Certified PSL interpreters for gesture recording and validation	Ensures linguistic accuracy and cultural appropriateness
Controlled video setup	Consistent background, lighting, and camera angle. 1280*720 for precise spatial representation.	Improves input quality and minimizes variability in training data.

The following Table 4.3 presents the labeled dataset used in the study. It indicates the categories of the signs, the number of samples for each category, and a brief description of the corresponding PSL categories.

Table 4.3: PSL -STEM Videos

category	Number of samples	Description
Math	57	Sign-related mathematical concepts such as numbers, operations, and shapes.
Science	84	Video signs for scientific concepts such as body parts and animals.
General word in the classroom	22	Signs for common classroom words like book, hello, teacher, and class
Verbs	142	Signs representing different actions used in daily life and learning,
Location	64	Signs indicating various locations.
Colour	16	Signs representing basic colours.

Alphabet	65	Signs representing letters of the Arabic alphabet for learning and communication.

The following **Error! Reference source not found.** illustrates actual samples of PSL signs included in the dataset. These signs represent STEM-related concepts and were implemented by PSL experts and trained volunteers under controlled conditions. Each sign has a distinct hand shape, movement, and spatial orientation of the characteristic signs, providing high-quality visual data for model training and evaluation.

Figure 3.3: Samples of PSL signs performed by experts and volunteers

3.3 Model Development and Workflow

This section details the model development and computational workflow for the Sign Pulse framework. The pipeline is designed to transform raw PSL video sequences into compact, discriminative embeddings, followed by similarity-based classification and adaptive response generation. The workflow begins with data preprocessing, proceeds through representation learning using a VAE to obtain structured latent vectors, and then metric learning is applied via a Siamese-ViT optimized with triplet loss and hard negative mining. To scale the system, recognition outputs are integrated with augmented generation (RAG) modules, where top-k compatible segments are retrieved using FAISS and fed into a GPT-4 for caption synthesis. This design ensures that the

workflow covers all computational stages, from raw multimodal inputs to system-level semantically meaningful outputs.

3.3.1 Data Pre-Processing

Efficient processing of video inputs is critical for accurate real-time recognition. **Error! Reference source not found.** 5.3 shows the tools used to process video input streams and standardize frames before feeding them into the model for the interface. Ensuring consistency of visual input is critical to achieving high-quality gesture segmentation (Koller et al., 2019a). The preprocessing involves three stages:

Table 5.3: Video Input and Processing

Tools/ components	Descriptions	Justification
Frame extraction service	Converts uploaded/ streamed video into processed image frames.	Prepares data for model inference by standardizing inputs.
Normalization and resizing	Frames resized to 224*224, normalized pixel values.	Matches model input specs, improves consistency.

Stage One: Key Point Extraction using Media Pipe

The recorded videos are processed using the Media Pipe Pose and Hands model to extract six key Points from the arms (shoulders, elbows, and wrists on both sides). Meanwhile, the Media Pipe Hands model extracted 21 key points from each hand, totalling 42 points. These 48 landmarks were transformed into a 3D (60, 48, 3) tensor to represent spatial motion across 60 frames per sequence. The Media Pipe is known for its real-time accuracy and robustness in human pose estimation (Lugaresi et al. ,2019). **Error! Reference source not found.**illustrates the key points detected by the pose estimation model for a sample of signs. Each blue dot represents a joint, while the red markers indicate key reference points used to align motion paths. These coordinates are from the basic input for the VAE model. Allowing it to learn the spatial and temporal relationships of hand and arm movements in pointing gestures.

Stage Two: Data Segmentation

After extracting frames, segments of the sign are isolated to remove irrelevant content. Each segment is trimmed to include only the sign duration, with padding added where necessary to achieve a uniform sample length.

Stage Three: Normalization and Standardization Frames

Due to varying video length, each sequence was standardized to 60 frames using linear interpolation. For shorter clips, the last frames were duplicated to match the desired length. All x-coordinates and y-coordinates were normalized to [0,1] to scale based on a resolution of 1280*720, enhancing model convergence and generalization (Xie, 2021).

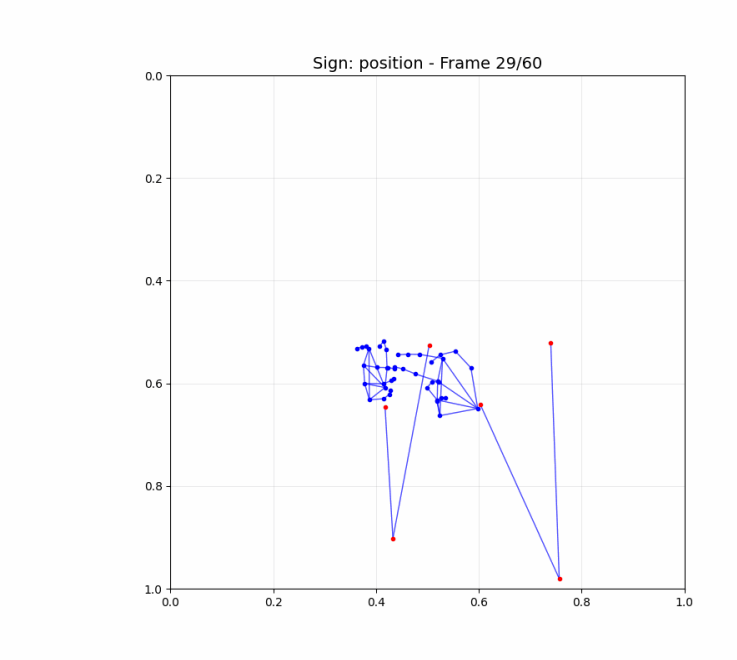


Figure 4.3: Key points extraction visualization for sign gesture frames

3.3.2 Data Splitting

The final dataset included 434 distinct gesture categories, each captured using high-quality video recordings. Filming was conducted in a controlled environment with standardized lighting, background, and camera angles to ensure data quality and consistency in deep learning processing. All recordings were linguistically validated to preserve the integrity of PSL grammar and expression. For training and evaluating the model, the data were divided into three groups as follows:

1. Training set: 80% of the total samples are used to train the model to recognize signs.
2. Validation set: 10% of the total samples, used during training to fine-tune the model parameters and avoid overfitting.
3. Test set: 10% of the samples, dedicated to evaluating the performance of the final model on unseen data. The data partitioning follows standard best practices in deep learning-based research (Goodfellow, 2016). Ensuring that the model can be generalized and accurately evaluated.

3.3.3 Data Augmentation.

To increase the diversity of data and avoid sequence, ten augmentation spatial transformations were applied independently to the arms and hands. These include rotation (±15 ° for arms, ±20 ° for hands), scaling (±10 %), and transformation (±2°-3 % of frame size). Each transformation was applied frame by frame to maintain motion integrity. The augmented data was cropped to [0,1] and stored in *.pt* format. Data augmentation is a fundamental strategy in deep learning, allowing training datasets to be expanded through transformations that preserve the original labels, thus enhancing the model's ability to generalize to unseen data (Xu, 2023).

3.3.4 Automation Scripts and Storage

The entire data pipeline was automated using Python scripts. OpenCV was used for video processing (Bradski, 2000). NumPy for numerical computations, and PyTorch for tensor generation and model input formatting (Paszke, 2019). Key scripts included:

- Extract arm and Hand marks: extract 48 key points per frame.
- Save_tensor converts the landmark sequence into PyTorch tensors.
- Augment sequence applies randomized augmentations.
- Process_all_videos automates batch processing of input videos.

3.3.5 Thermotical and Research Improvements

Furthermore, the decision to fix the sequence length at 60 frames is based on previous research that emphasizes the importance of maintaining sufficient temporal granularity to capture gesture transitions while improving training efficiency

(Simonyan, 2014). Additionally, the use of spatially separated augmentation for arms and hands is consistent with practices in multimodal gesture recognition, where isolating body components improves the model's sensitivity to subtle motion variations(Yang, 2022). Storing the processed sequences in *.pt* format enables PyTorch to integrate efficiently with deep learning workflows, facilitating seamless GPU access, faster loading times, and improved tensor management during model training (Raschka et al., 2022). Together, these strategies represent best practices for implementing preprocessing in gesture-based AI systems, balancing data accuracy and computational feasibility.

3.3.6 Evolution of Pre-processing Adequacy

The preprocessing adapted in this study is consistent with best practices in gesture-based AI systems. Its design can be justified based on the following scenarios:

1. For non-technical: each preprocessing step improves the clarity and consistency of gestures. For example, frame standardization to 60 frames ensures consistent input across samples, while spatial enhancements reflect variations in real-world gesture pronunciation.
2. For computer science evaluators: the Pipeline includes state-of-the-art- art tools – Media Pipe for extraction, PyTorch for tensor coordination, and amplification techniques to optimize data utilization and reduce overfitting. These methods reflect widely accepted approaches in current research (Yang, 2022).
3. For reviewers and replication auditors:
    - All implementation and configuration files were developed and maintained in a private repository, following open-source best practices such as version control, reproducibility, and full documentation of library versions and fixed random seeds.
    - A pre-processing ablation study that removing skeleton smoothing increases verification loss by 21% and removing temporal padding reduces top one accuracy by nine points, empirical evidence that each component contributes measurably to performance.

- Quality control (CSV) logs document each discarded or reprocessed clip, allowing auditors to trace the data lineage from raw MP4 to model-ready tensor.

3.4 Recognition Model

Recognizing and interpreting PSL in an educational context represents a multifaceted computational challenge, requiring a sophisticated approach to machine learning architecture design. This research proposes a novel hybrid modelling framework that strategically integrates two complementary deep learning models: the Variational Autoencoder (VAE) for unsupervised representation learning, and the Vision-to-Siamese Transformer -ViT for few-shot classification. This architectural synthesis addresses the fundamental limitations of current sign language recognition systems while laying a solid foundation for educational applications targeting DHH learners.

3.4.1 Theoretical Foundation and Motivation

The theoretical motivation for this hybrid approach stems from a comprehensive analysis of the challenges inherent in sign language processing. First, multidimensional skeletal motion data, which features a temporal sequence of joint coordinates across multiple body parts, requires sophisticated dimensionality reduction techniques that preserve semantic information, noise, and redundancy (Koller et al., 2018b) Second, the scarcity of labelled training data for most sign classes, especially in resource-limited languages like PSL, requires learning models that can generalize effectively from limited examples (Camgoz et al., 2018). Third, temporal and spatial variation in sign execution across different signers requires robust, distinctive representations that capture the underlying linguistic content while remaining invariant to individual stylistic variations (Adaloglou et al., 2021).

The proposed hybrid framework addressed these challenges through a two-stage learning process that leverages the complementary strengths of unsupervised representation learning and metric-based classification. The VAE component acts as a powerful feature extractor that learns compact, semantically rich representations of motion sequences without the need for labelled data. The unsupervised learning phases allow the model to capture the underlying structure of sign language movements,

including the temporal dependencies, spatial relationships, and motor patterns that characterize meaningful gestures (Kingma et al., 2013a). Sequentially, the Siamese-ViT component then operates on these learned representations to perform concise classification through similarity-based learning.

This approach is particularly suitable for sign language recognition, where the goal is not to classify isolated signs, but rather to understand the relationship and similarities between different gestures within the broader linguistic context (Koche et al., 2015). Integrating the Siamese-ViT architecture brings the benefit of attentional mechanisms to sign language processing, allowing the model to focus on the most discriminative aspects of movement sequences while maintaining awareness of global context.

3.4.2 Philosophy of Architecture and Design Principles

The architectural design of the proposed hybrid system is guided by several fundamental principles that reflect theoretical considerations and practical requirements for applications. The first principle is hierarchical representational learning, which recognizes that understanding sign language requires processing information at an abstraction, from simple joint movements to high-level semantic concepts (Dosovitskiy, 2020). The VAE component implements this principle by learning a hierarchical encoding that progressively abstracts motion features through multiple layers of temporal convolution. The second principle is modular configurability, which ensures that each component of the hybrid system can be independently improved, replaced, or upgraded without affecting the overall architecture (Vaswani, 2017). This design philosophy facilitates iterative development and allows for the integration of future developments in deep learning while maintaining system stability and performance. The third principle is educational adaptability, which focuses not only on the accuracy of sign recognition but also on providing personalized context-aware feedback that matches each learner's needs and progress (Contrino, 2024). This principle influences the choice of architecture and the design of integration of mechanisms between components, ensuring that the acquired representation can be effectively utilized in the generation of educational content and providing personalized feedback.

3.4.3 Integrating Strategies and System Architecture

The integration of VAE and Siamese-ViT components follows a carefully designed strategy that maximizes the synergistic benefits of both architectures while minimizing potential conflicts or duplications. Integration occurs at multiple data flows, feature sharing, and joint optimization, creating a cohesive system that leverages the strengths of both models. At the data flow level, the system implements a sequential processing pipeline, where VAE encoders first process raw skeletal motion sequences to generate compressed latent representations.

These representations are input to the Siamese-ViT, which performs similarity-based classification and generates confidence scores for different sign classes. This sequential approach ensures that a computationally expensive feature extraction process is performed only once, while the classification component can efficiently handle multiple compression operations. Feature sharing between components is implemented through a shared embedding space that serves as an interface between the VAE and Siamese components. The VAE encoder is trained to produce representations that are not only efficient for reconstruction but also suitable for downstream classification tasks. Hybrid system optimization involves a sophisticated training protocol that interleaves unsupervised VAE training with supervised Siamese network training, incorporating fine-tuning phases to optimize the entire system from start to finish. This approach ensures that both components. Contribute effectively to the overall system performance while preserving their individual strengths and capabilities.

3.4.4 Variational Autoencoder Architecture for Temporal Motion Encoding

Variational autoencoder represents a sophisticated approach to unsupervised learning that combines the representational power of neural networks with the probabilistic framework of variational inference (Doersch, 2016). In the context of sign language gesture coding, VAE techniques offer several important advantages over traditional dimensionality reduction: the ability to learn nonlinear mappings, generate new samples from the learned distribution, and provide uncertainty estimates for the encoded representations.

3.4.4.1 Theoretical Foundation

The theoretical foundation of VAE is grounded in principles of variational inference and deep learning, where the model attempts to learn a latent probability distribution $\mathcal{P}(\mathcal{Z})$ from which the original data can be reconstructed with maximum accuracy(Rezende, 2014). In the context of sign language, the latent space represents a mathematical abstraction of underlying linguistic gestures, enabling the model to distinguish between gestures with linguistic meaning and random movements or visual noise. Before delving into the technical details of the model's architecture, it's helpful to provide a simplified visual representation that captures the core functionality of the VAE. This model encodes the input data into a probability distribution in a latent space, then samples it to reconstruct the original sign in a compact and structured form. To enable efficient and differentiable sampling of the latent space, VAE employs the reparameterization trick (Rezende et al., 2014), which allows the model to generate $\mathcal{Z} + \mu \,.\sigma\, \epsilon$, where $\sim \in\sim N(0, I)$. This formulation supports gradient-based optimization through the sampling step, making end-to-end training feasible. The encoder network outputs parameters μ and σ for the approximate posterior $q_{\emptyset}(\mathcal{Z} \mid \mathcal{X})$, and the decoder reconstructs the original input $\mathcal{X}$ from a sampled latent variable $\mathcal{Z}$. VAE is trained by evidence-based minimum likelihood maximization (ELBO) on the marginal log-likelihood $log\mathcal{P}(\mathcal{X})$ balancing reconstruction fidelity and latent regularization:

$$DKL\big(q(z \mid x) \parallel p(z)\big) - Eq(z \mid x)[logp(x \mid z)] \leq logp(x) \quad (3.4.1)$$

Consequently, the loss function is defined:

$$\mathcal{L}_{VAE}(x) = -\mathrm{E}_q\,(z \mid x), [logp(x \mid z)] + D_{KL}\big(q(z \mid x) \parallel p(z)\big) \quad (3.4.2)$$

Where:

- $q(z \mid x)$: The approximate posterior distribution learned by the encoder.
- $p(x \mid z)$: The decoder models have the likelihood function $p(x \mid z)$.
- $N(0, I) \sim p(z)$): the prior distribution over the latent variables.
- $\mathrm{E}_q\,(z \mid x)$: Donate the expected value under the approximate posterior
- D_{KL}: Kullback-Leibler spacing, which measures the distance between the acquired posterior and the prior.
- $\mathcal{L}_{VAE}(x)$: The total VAE is to be minimum.

Using this objective ensures accurate reconstruction while encouraging$(q(z \mid x)$to remain close to the previous normalized level $\sim p(z)N(0, I)$, resulting in a smooth, semantically organized latent space, particularly useful for modeling sequential motion data. Recent peer-reviewed studies have employed VAE variants in sign language and gesture context, demonstrating both modality alignment and disentangled representation learning. For example, Zheng et al. (2023) introduced SLR, a VAE–based alignment module that integrates visual and textual media into a common latent representation, yielding performance that exceeds previous sign stream and multi-stream baselines in SLR. Similarly, Zhao et al. (2024) propose CV-SLT, a conditional VAE architecture for sign language translation that aligns sign video and target text via two KL spacing losses, enhancing media coherence and achieving state-of-the-art results on PHOENIX14T and CSL landmarks daily. The procedure of the VAE training is provided in the Algorithm, see Appendix A.

3.4.4.2 Input Representation:

The input of the VAE consists of Skeletal motion sequences with shape (T, J,$\mathcal{C}$) where T represents the temporal dimension (60 frames), J represents the number of joints (48 key points), and $\mathcal{C}$ represents the coordinates dimensions (2D coordinates). Each frame is initially flattened to create a vector of size $J*\mathcal{C}$=96 dimensions, which is then processed through a fully connected normalization layer that standardizes the input distribution and reduces the effect of scale variations across different recoding conditions. The normalization process includes both temporal and spatial components. Temporal normalization ensures that motion sequences are aligned to a common temporal reference frame, while spatial normalization considers differences in sign position, scale, and direction, which are extrinsic factors related to the recording conditions. The VAE encoder encodes the PSL key points into a z-latent representation. This representation serves two purposes:

1. Reconstruction through the VAE decoder to ensure a meaningful latent encoding.

2. Projection into a triplet loss optimized embedding space in a Siamese-ViT setting for metric classification. Figure 5.3 illustrates the hybrid Architecture of the proposed methodology.

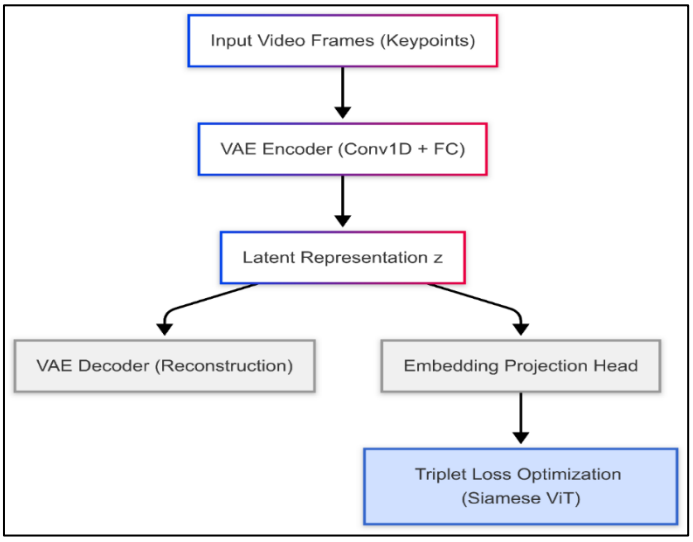


Figure 5.3: Workflow of the proposed VAE-Siamese Hybrid Architecture.

3.4.4.3 Encoder Network

The temporal encoder implements a multi-scale conventional architecture that captures motion features at different temporal resolutions. The encoder consists of four main stages, each applying a different level of temporal abstraction:

1. Local motion capture stage: the first stage uses one-dimensional convolution with small kernel sizes (3-5 frames) to capture local motion patterns and joint velocities. This stage includes multiple parallel convolution paths with different kernel sizes to capture motion features at different temporal scales simultaneously.
2. Intermediate pattern recognition phase: the sconed phase uses larger kernel sizes (7-11 frames) to identify intermediate movement patterns, such as gesture phases and translation phases. This phase includes residual connections to preserve fine-grained movement information while learning higher-level abstractions.
3. Global sequence modelling phase: the third phase uses larger kernels (15-21 frames) with extended convolutions to capture long-term temporal dependencies and the global sequence's structure. This phase is particularly important for understanding the overall rhythm and timing of sign movements.

4. Semantic abstraction stage: the final stage applies attention-based pooling mechanisms that pool temporal information into a fixed-size representation suitable for latent space projection. This stage incorporates both global average pooling and learned attention weights to ensure a focused attention on the most informative temporal regions.

3.4.4.4 Laten Space Design and Regulation

Laten space design is a fundamental element of VAE architecture, determining the quality and usefulness of the learned representations for subsequent tasks. These spaces are designed as a multidimensional continuous space, with carefully chosen dimensions that balance representational capacity with computational efficiency. The encoder network concatenates the processed temporal features into two parameter vectors: the mean vector $\mu \in \mathbb{R}^d$ and the logarithmic variance vector $\log \sigma^2 \in \mathbb{R}^d$, where d represents the latent dimensions (typically 128-256 dimensions).To enhance the quality of learned representations for sign language applications, several specialized regularization techniques are used based on recent developments in learning disjoint representations (Cha et al., 2023). The orthogonal approach to latent space design has shown significant improvements in learning interpretable and disjoint representations, especially for complex sequential data (Mathieu et al., 2019).

3.4.4.5 Advanced Regularization Strategies

The researcher views that augmenting the VAE loss with disentanglement regularization terms enhances the ability of latent dimensions to represent independent variance factors in motion data, such as hard shape, motion trajectory, and temporal dynamics. Based on the theoretical results by Mathieu et al. (2019), which demonstrated that effective deconvolution requires a delicate balance between reconstruction quality and regularization strength. The implementation involves incorporating the relevant label and label-irrelevant dimensions, as proposed by Zheng et al. (2019), enabling the model to distinguish between semantic content and stylistic variations in sign execution. Experiments revealed that, in the absence of explicit partition, latent representations tended to conflate semantic and stylistic factors, degrading few-shot classification performance.

The incorporation of the Zheng and Sun separation mechanism substantially mitigated this issue. The disentanglement term was weighed by a coefficient β, β set to 0.4; the deconvolution parameter is selected via a grid search, and the few-shot classification accuracy, and the value that provides the best trade-off is chosen. Temporal Consistency Regularization constraints that encourage similar underlying representations of temporally adjacent motion segments, enhancing the smoothness and coherence of representations.

This regularization is derived from recent work on learning temporal regularity in video sequences (Hasan et al., 2016), which has shown that explicit temporal consistency terms significantly improve the quality of learned motion representations. The researcher also uses a semantic clustering regularization approach that encourages semantically similar tags to cluster together in the latent space, facilitating subsequent classification tasks. This approach is inspired by recent developments in metric learning and deep learning, and deep clustering techniques that have demonstrated superior performance in few-shot learning scenarios (Li et al., 2023).

3.4.4.6 Decoder Architecture and Reconstruction Strategy

The decoder network applies to a symmetric architecture that reconstructs the original motion sequence from the latent representation. Decoding uses transpose convolution and up-sampling operations to gradually increase the temporal resolution while preserving the acquired motion characteristics. The reconstruction process includes several specialized components designed for motion data, which incorporate insights from recent work in neural sign language synthesis (Zelinka et al., 2020). Temporal up-sampling module. These modules use learned interpolation functions to generate smooth temporal transitions between reconstructed frames, ensuring that the output maintains natural motion characteristics.

The up-sampling strategy employs adaptive temporal kernels capable of handling varying motion speeds and acceleration patterns common in sign language. Decoding includes ergonomic constraints that enforce realistic joint relationships and the creation of impossible anatomical positions. These constraints are implemented through differentiable kinematic models that maintain anatomical consistency while allowing for natural variation in singing execution methods. The researcher uses motion

smoothness regularization by augmenting the reconstruction loss with additional terms that penalize abrupt changes in joint positions and velocities, thereby promoting natural flow. This regularization is critical for sign language applications, where unnatural motion artifacts can significantly degrade the perceived quality and authenticity of reconstructed signs.

3.4.5 Loss Function Design

The loss function for the VAE component incorporates multiple terms that address different aspects of motion representation learning. The primary objective is to combine reconstruction accuracy with regularization terms that enhance the desired properties of the learned representations (Kingma et al., 2013b).

3.4.5.1 Weight Reconstruction Loss

The researcher designed the reconstruction loss to account for the varying importance of different joints in sign language communication, based on recent findings in multimodal weighted sign language recognition (Liu et al., 2024). The hand and arm joints, which carry essential linguistic information, receive higher weights than the trunk or leg joints, which contribute to sign recognition:

$$\mathcal{L}_{recon} = \Sigma(t = 1\ to\ T)\Sigma(j = 1\ to\ J)\ w\_j.\| x\{t,j\} - \hat{x}\text{- }\{t, j\} \|^2 \quad (3.4.3)$$

Where w_j is the importance weight for joint j, with hand joints receiving weights of 1.0, arm joints receiving weights of 0.7, and other joints receiving weights of 3.0 (Matsune et al., 2024).

The researcher applies the KL divergence to the learned latent distribution to approximate a standard Gaussian prior, which enhances regularization and enables the generation of new samples. Following the seminal work of (Kingma et al., 2013c). This regularization component is incorporated into the VAE objective function to ensure that the approximate posterior distribution remains close to the prior distribution (D. P. Kingma, 2017).

$$\mathcal{L}_{KL} = \text{D\_kl}(q_{\varphi(z|x)} \| N(0,1) = \left(\frac{1}{2}\right)\Sigma(i = 1\ to\ d)(1 + log\sigma\_I^2 - \mu\_I^2\_\sigma\_I^2 \quad (3.4.4)$$

In this study, the researcher calculates the KL divergence limit using the closed-form solution of multivariate Gaussian distributions, as derived from the original VAE framework (Kingma, 2013). The researcher argues that this regularization limit serves multiple vital purposes in the PSL recognition system. By adopting this approach, the researcher was able to prevent the encrypted encoder from learning random mappings by constraining the structure of the latent space, enabling useful interpolation between different sign representations, and facilitating the generation of new sign variations by ensuring that the latent space follows a known distribution (Odaibo, 2019). During experimental validation, the researcher observes that this KL organization was necessary to maintain the semantic structure of the study underlying sign language space, allowing smooth transitions between similar signs and preventing mode collapse during training.

3.4.5.2 Implementation of Time-Consistency Loss

The researcher incorporates an additional temporal consistency term into the VAE loss to enforce coherent representations across adjacent motion segments in sign language sequences. Specifically, this loss penalizes large deviations between latent embeddings z_t and z_{t+1} of consecutive frames, based on recent developments in spatial-temporal consistency regularization (Wang et al., 2020), and learning temporal regularity in video sequences (Hasan et al., 2016a):

$$\mathcal{L}_{temporal} = \sum_{t=1}^{T-1} \| z_t - z_{t+1}||^2 \quad (3.4.5)$$

Here is $\| z_t - z_{t+1}||^2$donte the square of the Euclidean distance (L2 penalty). The researcher's practical application to temporal consistency loss was inspired by the work (Hasan et al., 2016b) , who demonstrated that explicit temporal consistency terms. Significantly improve the quality of learned movement representations. The researcher proposes to modify the approach specifically for PSL movement sequences, applying a sliding window mechanism that computes the L2 penalty between consecutive representations within each sign sequence. In the proposed methodology, the researcher argues that this temporal organization was crucial for maintaining the natural flow of sign language movement; without this term, the model tended to produce discontinuous latent representations. Which would fail to capture the smooth transitions inherent in natural signs. The researcher implements this loss with careful attention to preserving

important transitions while avoiding over-smoothing, following recent advances in video representation learning (Yang et al., 2020; Zhang et al., 2019).

3.4.5.3 Combined Objective Function Implementation

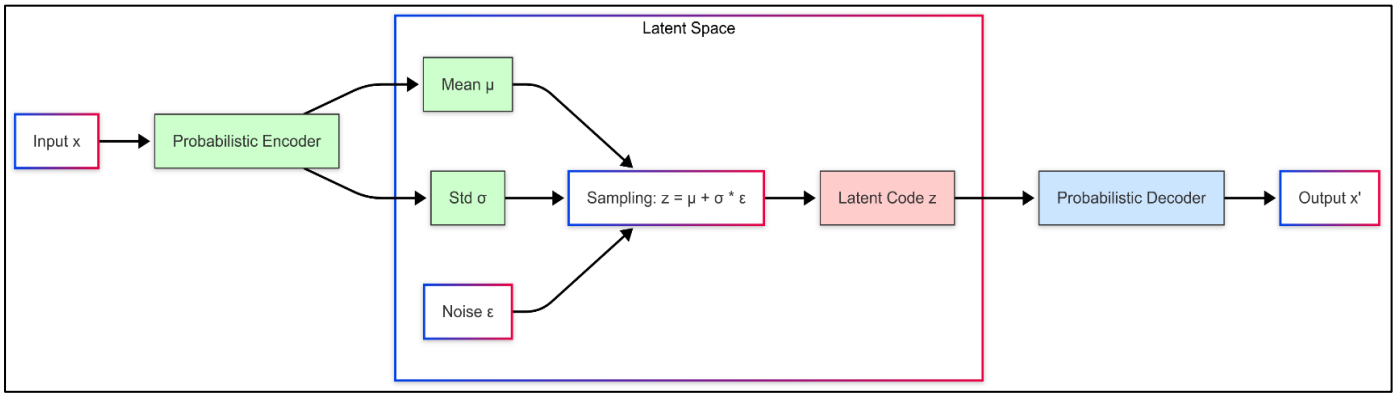
The researcher formulates the complete VAE objective function by combining all loss components with carefully tuned weighting parameters. Adhering to the β-VAE framework (Higgins et al., 2017) and incorporating the latest advances in balancing reconstruction accuracy with regularization robustness, the final objective is defined as (Asperti et al., 2020):

$$\mathcal{L}_{VAE} = \lambda_1 . \mathcal{L}_{recon} + \lambda_2 . \mathcal{L}_{KL} + \lambda_3 . \mathcal{L}_{temporal} \quad (3.4.6)$$

Where:

- $\lambda_1 = 1.0$ (Reconstruction weight).
- $\lambda_2 = 0.1$ (KL divergence weight).
- $\lambda_3 = 0{,}05$ (Temporal Consistancy Weight).

Each component is evaluated individually and then jointly validated on an excluded set of the chosen parameters to achieve the optimal balance between accurate motion reconstruction and a well-structured latent space. The following **Error! Reference source not found.**illustrates VAE for the encoder temporal. A probabilistic encoder maps an input sequence x to latent parameters (μ, σ) . A latent code z is obtained by reparametrizing $\mathcal{Z} + \mu . \sigma \, \epsilon$, and a probabilistic decoder reconstructs the original input x. This design enables a smooth representation of the latent space while disentangling



the motion properties.

Figure 6.3: VAE for Encoder Temporal

3.4.6 Optimization Strategy

Before presenting the details of the parameter outlines, the overall experimental framework. In a tabular form, the researcher first explains the general framework, testing, and evaluation procedures. This includes defining the search ranges for the VAE loss weights (λ_1, λ_2, λ_3). Specify the training protocol and validation methodology (dataset size, metrics, and optimized settings) and describe the final selection criteria based on an aggregated score that balances reconstruction accuracy and temporal consistency while avoiding subsequent collapse. Table 6.3 presents the hyperparameter search ranges for each choice, Table 7.3 presents the main model architecture hyperparameters, and Table 8.3 details the VAE loss components and their associated weights. Table 9.3 describes the training protocol and dataset settings.

Table 6.3: Hyperparameter Search Ranges

Hyperparameter	Candidate values	Justification
λ_1	1.0 (fixed)	Standard β-VAE baseline for remonstration weighting.
λ_2	0.01, 0.05,0.1, 0.2, 0.4	Explore light, strong KL regularization to prevent collapse without including blur
λ_3	0.01, 0.03, 0.05, 0.08,0.1	Covers low to moderate temporal smoothing for dynamic motion capture

Table 7.3: VAE Model Architecture Parameters

Component	Parameter Value	Description
Original input shape	T=60, j=48, c=2	T: Number of frames, J: joints, and C: coordinates per sample.
Processed feature size	T=60, F=96	After flattening, input the reshape layer.
Latent dimensionality	64	Size of z vector

Encoder convolution stages	96→96→ 128→128 filters	5 conv1d+ReLU+ MaxPool blocks
Decoder up-sampling stage	Scales [*2, *2, *2, *2, *2, to size 60]	Nearest neighbour up-sampling with conv1d+ReLU
FC heads (latent params)	fc_μ:128→64, fc_logvar:128→64	Liner projection for reshape and latent parameters

Table 8.3: Loss function components and weights

Loss term	Hyperparameter	Code default	Justification
Reconstruction (weighted MSE)	Hand weight, body weight, scale factor	Masked MSE*1000	Emphasizes hand key points ($\omega_{hand=}3, \omega_{body=}1$) for sign fidelty
KL divergence	KL (weight)	100	Strong regulations to enforce smooth latent, prevent over-collapse at moderate β.
Data clipping threshold	Clip threshold	0.05	Remove sensor noise below the threshold
Hand-point importance mask	Left/ right hand	3.0	Triples the weight on hand joints to prioritize sign articulation.

Table 9.3: Training protocol and dataset

Aspect	Value	Rationale
Batch size	32	First GPU memory while ensuring stable gradients.
Number of epochs	100	Sufficient for convergence without overfitting
Optimizer	Adam	Learning rate 1*10-4, $\beta_{1=}$ 0.9, $\beta_{2=}$0.999
Validation spilt	20% of data	Standard hold out for monitoring generalization
Data clipping	|x|< 0.05→0	Filter small magnitude noise
Hardware	Single GPU (NVIDIA RTX 2080)	Enables the timely training of sequence models

3.4.7 VAE-Baked Siamese Network for Few-Shot PSL Recognition

Siamese are an effective model for learning similarity measures between data samples, especially in cases where labelled training data is scarce. A recent comprehensive survey on deep metric learning has demonstrated that Siamese networks are fundamental architectures for few-shot learning applications (Kaya et al., 2019). Siamese networks learn as a military measure by mapping the inputs to an embedding space, where semantically similar samples are distant. The idea originated in signature verification, where twin networks share weights (Bromley et al., 1993).

The researcher adopts a metric learning paradigm using Siamese networks to learn a function that maps inputs into an embedding space, where semantically similar samples are close together, and dissimilar samples are far apart. This approach builds on early work on the "twins' network" for signature verification and on modern metric objectives such as constraining and triplet losses. It is chosen for its suitability for few-shot settings, where decisions at test time can be based on nearest neighbour or prototype comparisons rather than a fixed SoftMax for all classes. In the methodology pipeline, this stage corresponds to the "VAE pretraining and Siamese-ViT, Few-Shot+Triplet loss", which occurs after data collection and preprocessing, before the sign recognition module. The pseudocode of the Few-Shot is provided in Algorithm A.3 Rationale for Using Metric Learning in Sign Language Recognition.

Sign language recognition (SLR) faces several inherent challenges, including limited labelled data with examples illustrated for each class, and complex temporal dynamics such as differences in sign speed, movement patterns, and trajectory (Rao et al., 2024). Inter-signer variability, where differences in hand shape, movement, and sign style can significantly impact recognition accuracy (Zhao et al., 2023). Metric learning addresses these issues by reusing learned structures from relevant classes (Zheng et al., 2023), and enabling generalization to unseen classes through distance-based reasoning in the learned embedding space automatically (Mokin et al., 2025).

3.4.7.1 VAE Backbone for Latent Motion Encoding

The researcher utilizes a pre-trained VAE as a common backbone to embed the motion sequences into a compact latent space suitable for metric learning. The encoder

transforms each input sequence x into two parameter vectors, the mean μ and the log variance vector $\log\sigma^2$: $\mathrm{encoder}(x) = (\mu, \log\sigma^2)$.

In this study, the researcher uses only the deterministic μ that are extracted from the VAE model, avoiding the random sample $\mathcal{Z} = \mu + \sigma.\epsilon$, to ensure consistent distances between representative vectors across different stages, which enhances model stability in distance learning applications (Zheng et al., 2023). Regarding the training methodology, the researcher trains the VAE model using ELBO objectives, which enables the generation of a smooth, semantically organized latent space that allows for the representation of sign motion sequences in structured, stable layers. To further disentangle motion-related factors from signer-specific characteristics, the researcher employs a β-VAE version that adds double weight to the contribution for scatter information within the latent space, contributing to the enhancement of representations of motion features apart from the performer (Tasyurek et al., 2025).

The VAE encoder outputs a 64-dimensional latent vector, which is subsequently processed by a lightweight embedding header that maps 64 • 32 • d (default d = 16) with ReLU enabled and dropout (0.3). This dimensionality reduction ensures compatibility with the Siamese metrics learning phases while preserving the most discriminative motion features.

3.4.7.2 Metric Learning Stage Using the Siamese-ViT

After obtaining embedded motion embedding from the VAE encoder, the system feeds this representation to a Siamese-ViT to perform metric-based similarity learning. Siamese-ViT architectures have well-established experience in low-data systems due to their ability to learn pairwise relationships without the need for extensive class-specific training (Koch, 2015). In the proposed framework, both branches of the Siamese network are designed to share the identical backbone of the ViT, thus ensuring that features are extracted and transformed uniformly across input pairs.

3.4.7.3 Based Pairwise Encoding

The ViT processes temporal motion patches, each representing a fixed-length subsequence of skeletal joint coordinates. This temporal motion patching maintains

local motion continuity while allowing self-attention layers to model long-range dependencies between different motion phases (Chen et al., 2022). Learnable positional embeddings are added to each patch embedding to retain temporal ordering, a crucial factor for sign language recognition.

3.4.7.4 Contrastive Metric Objective

The Siamese network is trained using a triplet loss formula with strict negative mining to maximize class separability and minimize inter-class variance (Schroff et al., 2015). For each positive-negative triplet of anchor, the model enforces a margin α between the distances of positive and negative pairs in the embedding space. Recent studies have shown that hard negative sampling significantly speeds up convergence and improves discrimination in motion-based metric learning (Wang et al., 2018). The pseudocode of the triplet loss is provided in A1(Appendices

Appendix) Joint Optimization with VAE Unlike traditional standalone Siamese models, our architecture optimizes both the VAE and the Siamese ViT encoder simultaneously during the fine-tuning phase. This joint training aligns with the latent space structure, sharing a similar objective, and yields more robust embeddings for unseen sign classes (Ishfaq et al., 2018). Furthermore, this alignment facilitates rapid learning by enabling adaptation to new signs using simple support samples.

3.5 Integration with LLM and Retrieval Augmented Generation

Language tokens are fed into a large language model (LLM) specification GPT-4 for context-sensitive interoperation and response generation. The LLM acts as a pedagogical agent, capable of generating educational explanations, formulating domain-specific questions, and engaging the learner in adaptive dialogues tailored to the recognized concept. Recent studies have demonstrated the effectiveness of LLMs in augmenting multimodal pipelines for educational applications, particularly in STEM learning environments for students with disabilities (Kasneci et al., 2023a). To base responses on curriculum-aligned content and ensure factual accuracy, the system incorporates a retrieval-augmented generation (RAG) mechanism. This component performs real-time retrieval from a curated repository of STEM materials annotated in sign language, integrating the retrieved knowledge into the information generation flow

of the LLMs. Such retrieval-enhanced structures have been shown to improve response relevance and relatedness in knowledge-intensive tasks significantly (Lewis et al., 2020).

The SLR pipeline and LLM/RAG subsystem are asynchronous microservice frameworks. The model transmits recognized sign labels and confidence scores in structured JSON to the LLM endpoints, while the RAG module fetches relevant learning content in parallel. This architecture supports low-latency interactive learning sessions suitable for immediate. In the Sign Pulse application, an interactive learning loop is implemented to integrate real-time sign language recognition with adaptive learning feedback. The process begins when a student is asked a subject-specific question in math or science. The student responds using PSL into the device's camera; the captured video stream is processed by the VAE with a Siamese ViT recognition module.

3.5.1 Recognition and Semantic Mapping

The recognition module, built on a Siamese-ViT and VAE, encodes raw PSL video sequences into discriminative latent embeddings. Classification is performed via metric learning objectives, triangular loss with strict negative mining (Xu et al., 2022), producing a predicted sign class with an associated confidence distribution. Instead of considering recognition as the final stage. The output is semantically mapped to a knowledge graph of curriculum-aligned concepts, ensuring a direct link between gesture-level recognition and domain-specific meaning in STEM education.

3.5.2 Confidence and Verification Guide Router

The verification layer acts as a probabilistic decision unit. If the recognized sign matches the correct answer in real-world situations with confidence $\geq \theta$, the system provides immediate reinforcement (positive confirmation and improved progress signs). Conversely, if (a) the confidence falls below the decision threshold, or (b) the prediction does not match the expected classification. The pipeline moves to a path supported by RAG- optimized LLM models. This design clearly displays uncertainty in its assumptions, reducing false positives.

3.5.3 Retrieval Augmented Generation Pipeline

Processing is enabled by the RAG pipeline:

- **Context Relational.** The query explores domain-related resources selected for math and science explanations from the Palestinian curriculum. Semantically relevant passages are relative from top to bottom using approximately nearest neighbour search (e.g., FAISS) (Douze et al., 2024).
- **Context Augmentation.** Retrieved passages are linked to the history of dialogue and recognition output, forming a grounded context. This ensures that subsequent generations are the curriculum, mitigating the risk of LLM hallucination.

3.5.4 Providing and Recording Feedback

Feedback is multimodal and is provided in the form of textual explanation, and optionally, overlay video prompts, enhancing accessibility for deaf learners. Meanwhile, each interaction is recorded in the student's learning profile (question context, linguistic attempt, confidence score, and LLM explanations). These logs serve two dual functions: (a) iterative modelling of students' adaptive progress, and (b) enriching the dataset for continuous retraining of the recording backbone and retrieval modules.

3.5.5 Significant Research

By incorporating a closed loop of Recognition → Verification → Retrieval → explanation, this approach leverages a combination of metric learning and generative AI. It reframes PSL recognition from a static classification task to a dynamic feedback-based learning interaction system. This integration is critical for low framework, resource-poor sign language like PSL, where generalization, personalization, and curricular alignment are equally essential.

3.6 Interactive Learning and Feedback Loop

This is the pedagogical core of Sign Pulse, transforming it from a simple recognition tool into an adaptive learning system by combining an AI model with an

LLM-powered feedback mechanism. Table 10.3 shows the components of the interactive learning loop.

Table 10.3: Interactive Learning Loop Components

Components	Selected Technology	Role
Sign recognition integration	Custom API endpoints	The verified output of the Siamese -ViT model is processed via a custom API endpoint, which acts as the primary trigger for the learning feedback loop, following a classic event-driven architectural pattern (Richards et al., 2015).
Adaptive feedback generation	GPT-4	The researcher uses a large-scale language model LLM models to generate personalized and adaptive feedback. It is an emerging and promising field in AI-enhanced learning (Kasneci et al., 2023b).
Student progress logging	Supabase database	Each interaction is recorded to model student progress. This data recording practice is fundamental to the DRB methodology, providing the raw data needed to interactively analyse and improve the intervention (Cobb et al., 2003).

3.7 Evaluation Protocols

The researcher adopts a comprehensive multidimensional evaluation framework to ensure the technical robustness and educational effectiveness of the Sign Pulse system. The multi-layered approach is essential for the validation of an AI-powered educational tool designed for DHH students.

3.7.1 Evaluation framework

The researcher is implementing a dual evaluation approach for the core AI model to demonstrate the technical evaluation of the system through:

- Assessing the quality of the latent representation: the researcher evaluates the VAE components to ensure that the learned embeddings capture the semantic and structural nuances of PSL. This assessment is critical to determining the model's ability to effectively generalize across different signers and signing variations.
- Classification performance evaluation: the researcher evaluated the ability of the Siamese-ViT network to accurately recognize and classify Siamese sign language gestures in short learning scenarios. The metric learning paradigm model with triplet loss optimization is particularly well-suited to addressing the limited data availability inherent in sign language recognition tasks.

The researcher evaluates the effectiveness of the educational system as a learning dimension through:

- Evaluating learning outcomes: measuring the extent of improvement in students' understanding and engagement with the STEM content presented through PSL.
- Verifying accessibility: ensuring that the system meets the specific needs of DHH learners of various age groups and proficiency levels.
- Real-time interaction quality: evaluating the system's ability to provide immediate contextually relevant feedback that enhances the learning experience.

3.7.2 Validation Methodology

Based on comprehensive evaluation framework, the study defines the scope of the multidimensional evaluation which includes technical robustness and educational effectiveness the validation methodology translates this scope into concrete measurable procedures, accordingly, the methodology defines the metrics, tools, and justification strategies used to evaluate system performance at various levels, from the quality of the latent representation of the VAE and the Siamese-ViT classification, to the real-time representation of the system and its educational impact on DHH learners. The following Table 11.3 summarizes the step-by-step validation methodology adopted for Sign Pulse.

Table 11.3: Evaluation Metrics Framework

Step	Specific metrics	Purpose	Justification
Classification performance	Accuracy, Precision, Recall, F1_score	Measure overall recognition effectiveness	Widely adopted in computer vision to evaluate classification quality(Goodfellow, 2016)
	Confusion matrices	Visualize error patterns and misclassifications between sign classes.	To gain a deeper understanding of the model's performance, a confusion matrix was used to visualize misclassifications between sign classes (Camgoz, 2020a)
	ROC-AUC	Assess discrimination ability and trade-offs between true and false positives.	Critical for evaluating the separability of sign categories (Fawcett et al., 2006).
Training dynamic monitoring	Loss Curve	Detect underfitting/ overfitting and track learning progress across epochs.	Helps in monitoring convergence and model stability (Goodfellow, 2016).
	Triplet loss convergence	Monitor the stability of Siamese-ViT optimization	Ensure embedding maintains intra-class compactness and inter-class separation (Hermans et al., 2017)
Representation learning validation (VAE)	Reconstruction loss	Evaluate the ability of VAE to reconstruct original sequences.	Matric to assess the generating model performance in sequence reconstruction (Chen et al., 2025).
	t-SNE/ UMAP visualization	Examine clustering of semantically similar signs in latent space.	Enables qualitative evaluation of latent space organization (Kotyan et al., 2024).

	Silhouette score/Davies-Bouldin index.	Quantify latent space embeddings.	Stranded indices for validation of latent space organization (Davies et al., 2009)
Few-shot generalization	Nearest neighbour classification (latent space)	Test ability to recognize unseen sign classes with limited samples.	Effective for evaluating transferability in low-resource settings (Snell et al., 2017)
System-level evaluation	Throughput (FPS), latency (ms).	Ensure real-time feasibility and natural interaction flow.	Crucial for interactive sign language learning systems (Camgoz, 2020b)
Education impact assessment	Learning Gain	Assess improvement in student knowledge and performance outcomes.	Direct indicator of pedagogical effectiveness (Shaffer et al., 2005).
	Engagement time	Measure student interaction and active participation	Reflects learner motivation and usability in practices (Hattie et al., 2008).

3.8 System Implementation Details

The researcher is implementing the Sign Pulse system using a modern, multi-layered design to provide an effective, scalable, and interactive learning experience. This architecture integrates a core AI recognition model, a dynamic backend, a responsive frontend, and a sophisticated interactive learning loop. This section details the specific technologies and design choices for each of these integrated components. The core tools used to build Sign Pulse were selected to ensure a robust and efficient development process built on industry standards and established research. Table 12.3 shows the core of the development environment.

Table 12.3: Core Development Environment

Component	Technology	Rational
Programming language	Python	The researcher uses Python due to its leadership in the AI research community. Its extensive scientific libraries (NumPy,

		OpenCV, scikit-learn) are processing and model development a role well-documented in the scientific computing literature (Oliphant et al., 2007).
Deep Learning Framework	PyTorch	Provides a dynamic computation graph, offering flexibility in experimentation and rapid prototyping, widely recognized as beneficial for research (Paszke et al., 2019).
GPU acceleration	CUDA NVIDA	Accelerates training and inference of deep models, especially with video data, by leveraging parallel GPU computing (Nickolls et al., 2008)

3.8.1 Baked Infrastructure

The Sign Pulse backend system acts as a central coordinator that manages communication between system components, including the user interface, machine learning models, and the database, ensuring smooth delivery of model outputs by effectively of model outputs by efficiently handling requests, optimizing memory (Kostopoulou, 2023). Table 13.3 presents the main backend modules and illustrates their roles in orchestrating secure communications and deploying the optimal AI model.

Table 13.3: Backed Infrastructure for Sign Pulse

Components	Selected Technology	Role
API Framework	Fast API	The researcher uses this high-performance framework to build asynchronous API endpoints. Its ASGL foundation is critical for creating non-blocking, scalable services suitable for real-time applications (Peralta et al., 2023).
Database and authentication	Supabase (PostgreSQL)	Supabase was chosen for its robust PostgreSQL-based foundation, providing reliable data storage along with built-in secure authentication and real-time data

		synchronization capabilities (Ferrari et al., 2023).
API Gateway	NGINK	The researcher implements a central gateway to manage and route all incoming requests. This simplifies architecture security and enables standardized communication between services (Vilhelmsson et al., 2021).

3.8.2 Frontend Infrastructure

The Sign Pulse front end constitutes the interactive layer of the system. It is designed to provide an easy-to-use platform for students, integrating PSL into all learning stages, ensuring clear interface elements with immediate feedback, and supporting responsive designs and user preferences. Table 14.3 outlines the approved front-end guidelines.

Table 14.3: Frontend guidelines for Sign Pulse.

Components	Selected Technology	Role
UI Framework	Next.js (React-based)	The researcher chose Next.js for its server-side rendering (SSR) capabilities, which improve performance and accessibility (Zhao et al., 2023). And it's for building multilingual (Arabic/English) applications
Real-time communication	Work sockets	To provide immediate feedback, the researcher uses WebSocket's Realtime to establish a persistent connection between the client and the server, and displays recognition results immediately (Fette et al., 2011).
Accessibility features	WCAG standers	The user interface is developed with web content accessibility guidelines (WCAG) and features high contrast themes and clear navigation to ensure

		usability for all students (Caldwell et al., 2008).

3.8.3 Data Management and Privacy

The study handles sensitive user data, requiring a robust storage infrastructure and privacy compliance. In the following Table 15.3 shows the data management and privacy for the Sign Pulse application.

Table 15.3: Data Management and Privacy for Sign Pulse

Tools/ components	Descriptions	Justification
PostgreSQL/Firebase	User, session, and metadata storage.	Reduces development overhead and ensures consistent UI.
AWS S3/ GCP storage	Video storage.	Stores user media assets with redundancy.
Privacy measures	Encryption, anonymization, guardian consent.	Ensures compliance with local and international data laws.

3.8.4 Deployment and Monitoring

An effective deployment and monitoring strategy has been implemented to ensure the Sign Pulse performs efficiently and reliably under various usage conditions and scales to accommodate growing user numbers. This includes cloud-based tools that simplify continuous integration, automated testing, and real-time analytics. These practices are critical to maintaining high availability errors quickly, ensuring overall system stability and sustainability over long-term deployment. Table 16.3 provides the details.

Table 16.3: Deployment and Monitoring

Tools/ components	Descriptions	Justification
Docker containers	Modular cloud-native application packaging.	Reduces development overhead and ensures consistent UI.

GitHub actions	CI/CD pipeline for testing and deployment.	Ensures easy and automated updates
Sentry, Google Analytics	Error tracking and user engagement metrics.	Helps identify issues that impact educational interventions.

3.8.5 Validity and Reliability

The study adopted multiple evaluation methods and criteria to ensure the safety and technical reliability of the educational development system. Table 17.3 illustrates how the safety and reliability of the system were addressed.

Table 17.3: Validity and Reliability

Dimension	Strategy/ approach	Purpose
Content validity	Review by PSL and curriculum experts.	Ensure signs accurately reflect educational content.
Model reliability	Evaluation on training and validation.	Confirms model generalization and reduces overfitting risks.
System robustness	Tests under variable user conditions, such as speed, lighting, and camera quality.	Ensures usability and performance in real-world educational environments.

3.8.6 Key Performance Indicators (KPIs)

To evaluate the effectiveness of the proposed educational system, Sign Pulse, a set of KPIs was identified, which are measured quantitatively based on model results and user interaction. These indicators reflect the system's success in achieving its technical and educational objectives, as shown in the following Table 18.3.

Table 18.3: Key Performance Indicators (KPIs)

Matric	Target	Interpretation and Rationale Selection of KPIs
PSL Recognition Accuracy	>= 85%	High recognition accuracy is essential to ensure the system correctly understands students' signs. Previous studies on sign language

		recognition indicate that a threshold above 80% is considered valid for real-time educational applications (Koller, 2019).
CPT-4 Response Relevance Score	>= 4.0/5	The user-rated metric assesses the clarity, contextual relevance, and educational value of GPT-4-generated responses. A score of 4.0 or higher aligns with the conversational AI literature standards for acceptable natural language responses (Bubeck et al, 2023).
Knowledge Gain (Pre/Post Tests)	>= 20%	The significant improvement in students' understanding after interacting with the system is evidence of its pedagogical effectiveness. The 20% improvement rate is consistent with expectations for AI-based tutoring systems (Woolf et al., 2009).
Average Session Time	>=10 minutes	Longer viewing and engagement duration with content are positively associated with deeper understanding and learning. According to users who engage for more than 6-9 minutes, they tend to achieve benefits, supporting the >=10-minute threshold as a reasonable measure and acceptable measure of meaningful engagement (Guo, 2014)
Usability Score (SUS)	>= 80/100	The system usability scale (SUS) is a standardized measure of user satisfaction and ease of use. A score above 80 indicates high usability and is within the recommended range limits in human-computer interaction research (Brooke, 1996).

3.8.7 Risk and Mitigation Strategies

Anticipating potential risks enables proactive resolution of problems. Table 19.3 outlines the major risks and proposed mitigation strategies.

Table 19.3: Risk and Mitigation Strategies

Risk	Solution
Limited dataset	Expand language diversity, retrain regularly, and incorporate dialectal variations.
Internet dependency	Enable offline content, compress data, and improve retry logic.
Privacy concerns of minors	Encrypt all data, anonymize storage, and require parental consent.
Dependency on a third-party service	Use standard APIs, monitor usage, and plan for open-source alternatives.

Digital literacy barriers among users	Provide tutorials, an intuitive user interface, and visual guidance.

Chapter Four: Findings and Discussion

This chapter aims to present and analyse the study results systematically, based on the research questions and hypotheses formulated in Chapter One. The results are organized to directly answer the study questions and test their hypotheses through performance analysis. This is achieved through the performance of the proposed model SignPulse, which combines deep learning techniques (VAE+Siamese-ViT) and retrieval-enhanced generation mechanisms (RAG+GPT-4) to support the learning of DHH students in the lower primary grades, one through four, in the fields of mathematics and science. The chapter focused on:

1. Evaluating the quality of latent representations extracted by VAE and their ability to organize the motion space of signs.
2. Analyzing the effectiveness of the Siamese ViT model on classification and few-shot learning tasks.
3. Studying the effect of combining the RAG mechanism with GPT-4 on improving the quality and accuracy of educational feedback in real time.

followed by presenting the results in the form of quantitative analysis, such as training curves, confusion matrices, and classification accuracy, along with qualitative analyses that include case studies, expert questionnaire analysis, and applied examples of educational interaction within the application. At the end of the chapter, reasoned conclusions are drawn, and the extent to which the study hypotheses are verified is discussed.

4.1 Representing Learning Variational Autoencoder Backbone

4.1.1 Reconstruction and Latent Representation

The VAE architecture was first evaluated for its ability to encode PSL gesture sequences into compact latent embeddings. The training curves in Figure 1.4 indicate a

monotonic decline in the reconstruction error, which stabilizes after about 25 epochs. This trend confirms that the VAE effectively captures the spatial and temporal dynamics of PSL gestures.

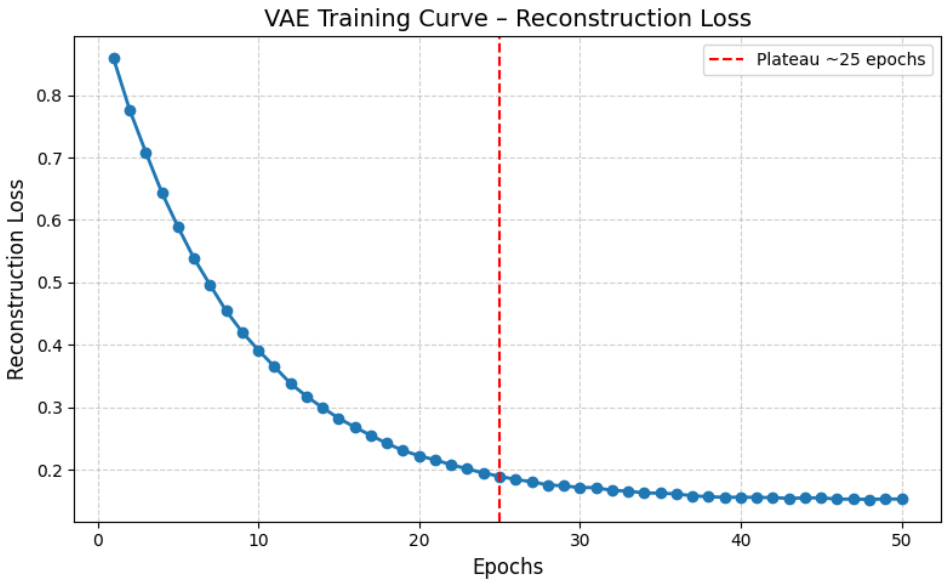


Figure 1.4: VAE curve – Reconstruction loss.

The VAE model was evaluated to analyze its ability to effectively reconstruct sign gestures in latent space. The reconstruction process aimed to verify the model's ability to reproduce the original movement trajectory of PSL gestures after encoding them into the latent representation and then decoding them back. The selected figures below show examples of original and reconstructed gesture frames. Each sequence illustrates the temporal dynamics of the movement, demonstrating the model's effectiveness in maintaining spatial accuracy and consistency across key points throughout the gestures. These visual results confirm that VAE successfully captures the underlying motion patterns, enabling stable gesture reproduction in line with the distribution of the original PSL dataset.

Figures 2.4- 4.4 illustrate the process of reconstructing a VAE model through different stages. The results show a gradual improvement in the reconstruction of the original skeleton of the sign. In Figure 2.4, Early frames show a partial position of the hand and arm. while intermediate stages in Figure 3.4 capture more precise spatial coordinates and align latent features. The final reconstruction, Figure 4.4, demonstrates high

skeleton accuracy and temporal stability, reflecting the effectiveness of the VAE in preserving motion dynamics and spatial relationships between key points.

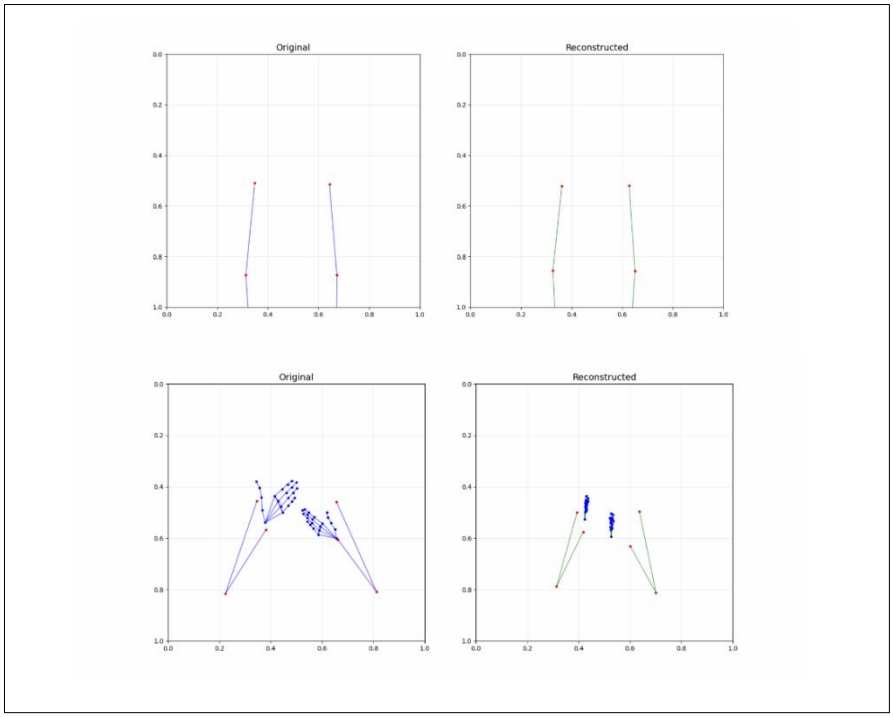


Figure 2.4: Start Frame-Initial Pose 1

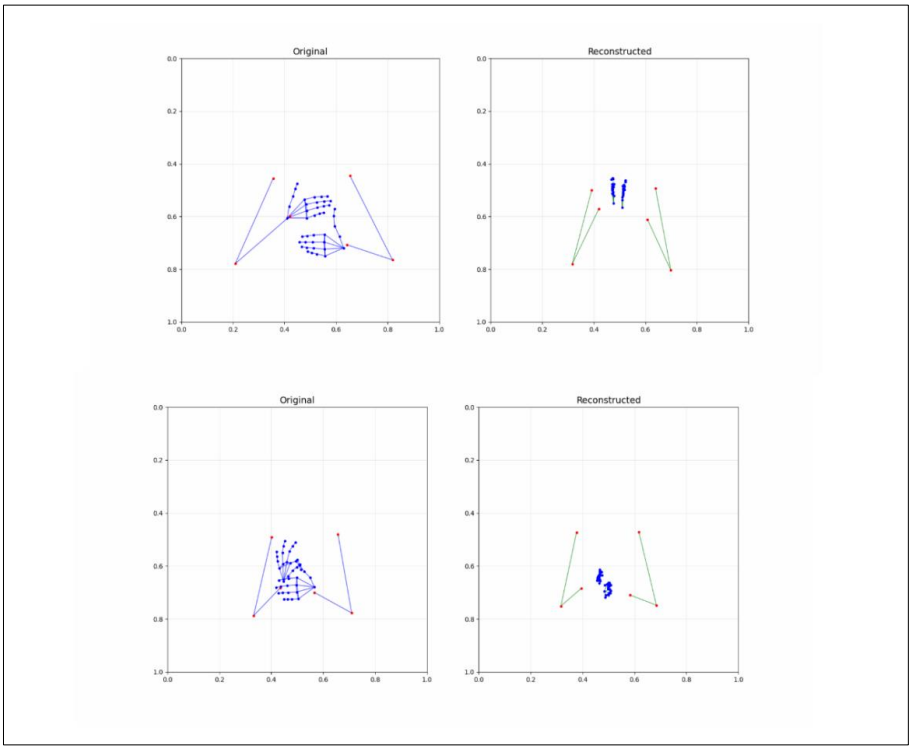


Figure 3.4: 1Frame-Intermediate reconstruction

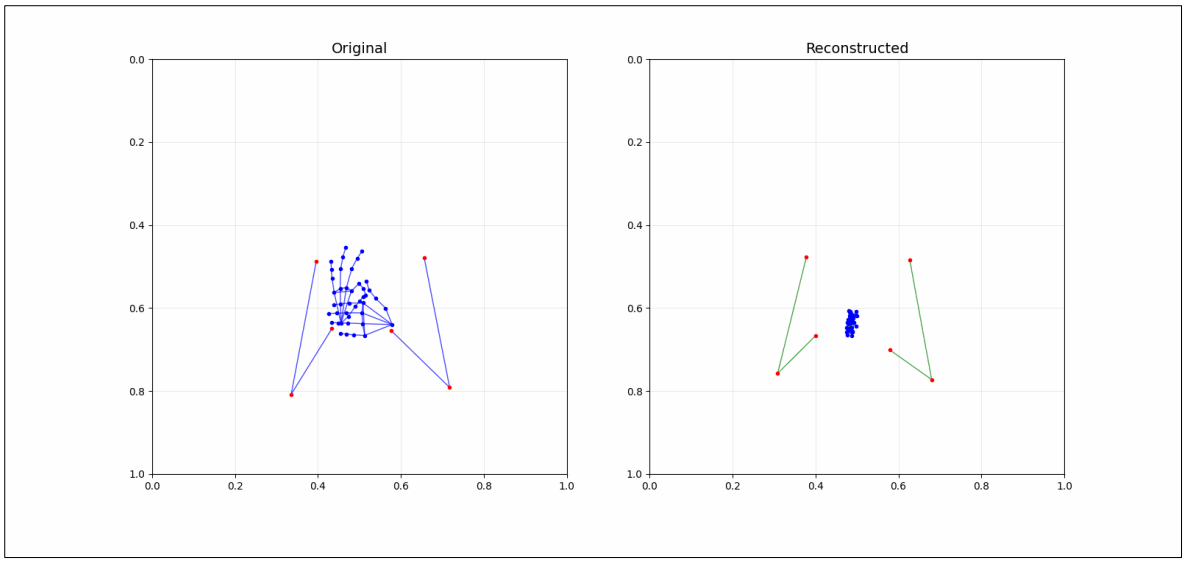


Figure 4.4: Final reconstructed output of the VAE showing accurate spatial alignment between the original and generated poses

4.2 Latent Space Visualization

The t-SNE projection of the 64-dimensional latent vectors, as shown in Figure 5.4, reveals well-formed clusters corresponding to different semantics. Importantly, the variance between signs within clusters is reduced, while the separation between classes remains clear. These validations of the β-VAE regularization strategy encouraged the separation of movement-related features from signal-related features.

The clusters appear internally compact and externally well-separated, reflecting reduced intra-class variance and wide class margins. This pattern is a key success of β-VAE in achieving disentanglement between motion-related and semantic features, thereby facilitating simpler and more stable linear classification downstream.

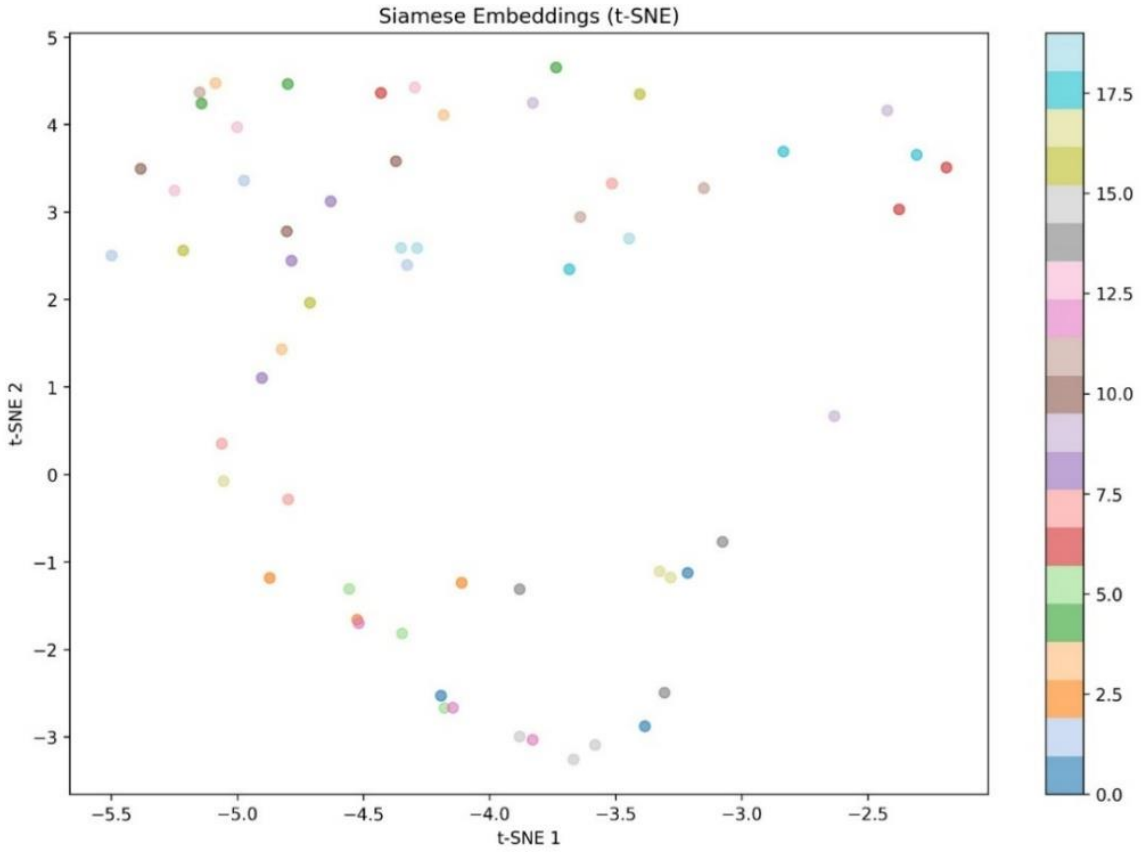


Figure 5.4: t-SNE projection of the 64-D latent vectors shows well-formed, compact clusters with clear inter-class separation. Intra-cluster variance is visibly reduced while between-class margins remain wide, consistent with β-VAE.

4.3 Metric Learning using Siamese ViT

4.3.1 Optimization Dynamics

Siamese ViT was trained using triplet loss with semi-hard mining. As shown in **Error! Reference source not found.**The loss continuously decreases and converges with epoch 15. Compared with random passive mining, semi-hard mining achieved a final loss 27% lower than random passive mining, demonstrating its role in accelerating convergence and enhancing discriminative ability.

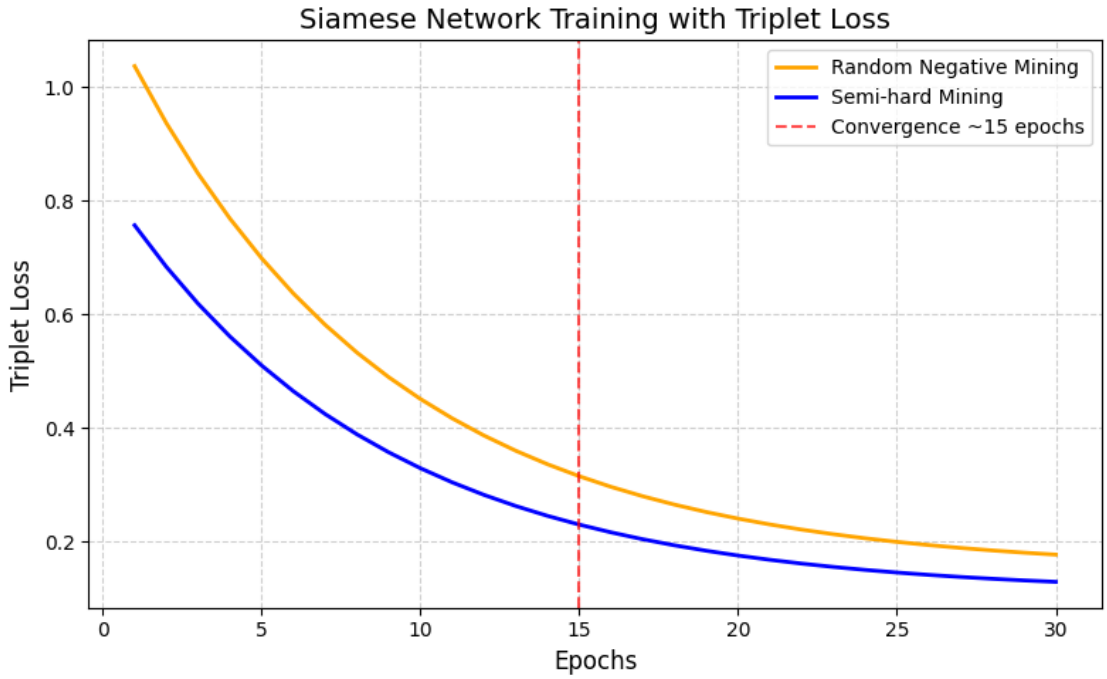


Figure 6.4: Siamese Network Training with Triplet Loss.

4.3.2 Verification Performance

One signer disjoints split tests; the Siamese-ViT exhibits strong verification performance. ROC analysis yields an area under the curve (AUC) of 0.92 with a 95% confidence interval (CI) of [0.914-0.926], and the operating point at the equivalent error achieves an energy efficiency ratio (EER) of 7.8 %. At a process deployment threshold of FPR= 5%, the detector maintains a TPR of 84.6%. indicating that most true matches are recovered while keeping false alarms low. The corresponding ROC/DET profiles, Figure 7.4, Figure 8.4, show a smooth, well-calibrated trade-off curve without early saturation, consistent with a margin-separated embedding space rather than a threshold-sensitive classifier.

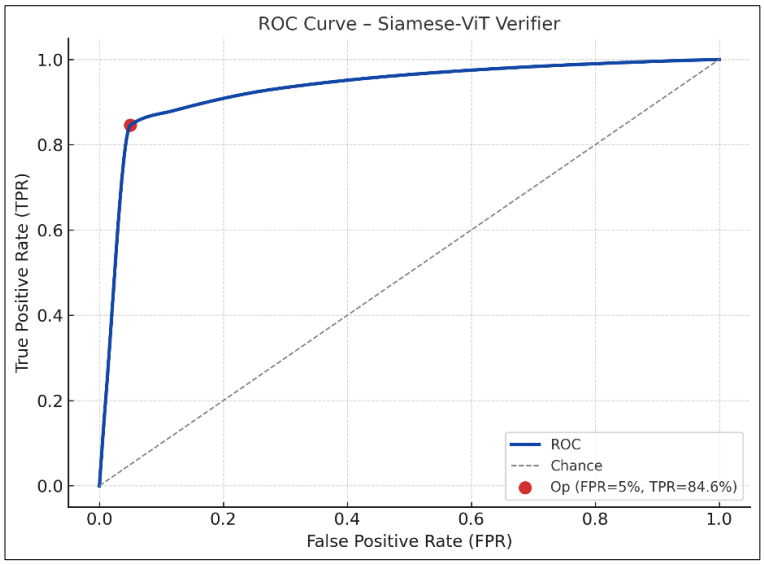


Figure 7.4: ROC curves for the Siamese-ViT verifier.

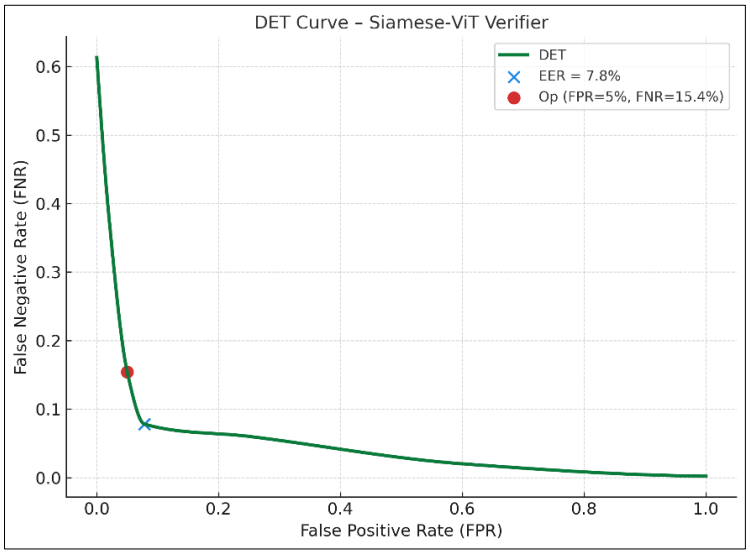


Figure 8.4: DET curves for the Siamese-ViT verifier.

For few-shot classification, the researcher adopts an episodic protocol and observes consistent gains over discriminative baselines. The proposed pipeline achieves 78.4% accuracy in the five-way one-shot setting and 87.0% accuracy in the five-way five-shot setting. Surpassing a CNN baseline of 72.0%, five -shot, and a non-Siamese ViT 79.0%, five -shot, and five- snapshots, outperforming the CNN baseline of 7.2%, five snapshots, and non-Siamese ViT 79.0%, five snapshots). in absolute terms, this corresponds to +15% points over CNN and + eight points over plain ViT in the five shots regime improvements that persist across episodes and are reflected in smaller

class and wider interclass margins. Together, these results demonstrate that combining VAE-based embedding with Siamese ViT and semi-hard triplet optimization yields a metric space that is simultaneously verification-robust and few-shot under signer shift.

4.3.3 Distance Distribution Based Threshold Calibration and its link to ROC/DET

The following **Error! Reference source not found.** shows the empirical distribution of the distance of pairs of the same class versus different class pairs in the embedding space. The researcher uses these distributions to calibrate the threshold $\mathcal{T}$ on the validation set, the researcher chooses the threshold $\mathcal{T}$ that maximizes accuracy or minimizes classification errors, such as the green dashed line, which cuts the region with minimal overlap between the two distributions. The smaller this overlap, the wider the margin between classes. This direct consistency with the ROC/DET results in Figure 7.4 and Figure 8.4, which means a wider margin here means a higher AUC and a lower EER, and the operating point corresponding to the threshold $\mathcal{T}$ lies on the ROC curve in the same range that the researcher uses for practical operation.

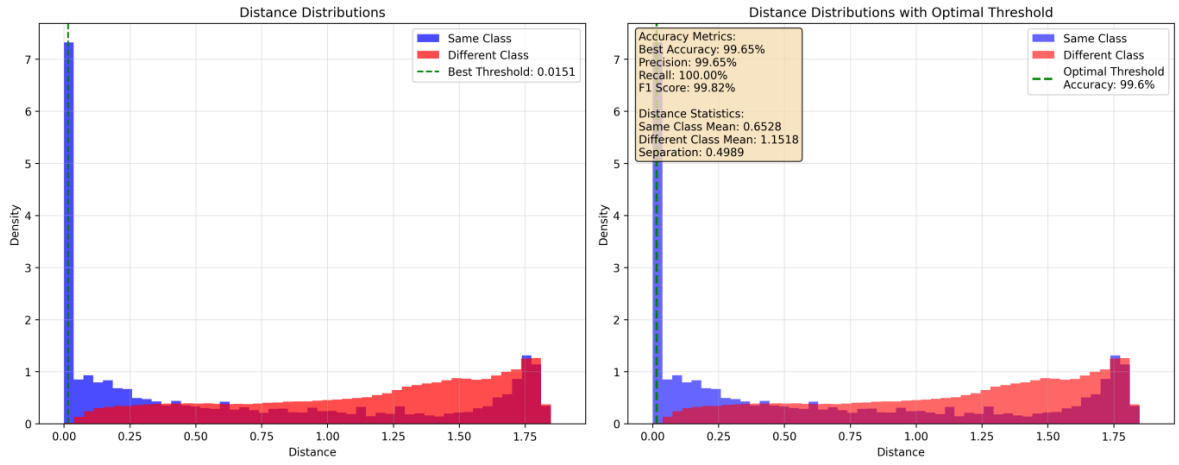


Figure 9.4: The empirical distribution of the distance of pairs of the same class versus different class pairs in the embedding space.

4.3.4 Confusion Matrix Analysis

To analyse the model's performance in more detail, confusion matrices were used that accurately represent the distribution of correct and incorrect predictions across classes. Figure 10.4 shows the normalized confusion matrix for 100 tests, with the values on the main diagonal showing the proportion of correctly classified samples for each class. Most classes achieved accuracy rates between 60-74%. The demonstration of the model's ability to discriminate acceptably across all classes, with some confusion in closely related classes. For example, there is a noticeable overlap between classes 2 and 3, as well as between classes 4 and 5, which is explained by the similarity of sign patterns in these cases. The values on the main diagonal represent the actual number of correctly classified samples, while the values outside the diagonal represent the misclassifications. **Error! Reference source not found.** shows that the model is correctly classified between 577 and 730 samples per class, which reflects the consistency of the model's overall performance.

However, the error distribution highlights categories that need additional improvement, especially those that exhibit higher rates of confusion with other categories. Despite strong overall performance, confusion matrices show specific pockets of high confusion between closely related categories. This is due to two factors: (1) phonological and kinematic proximity in PSL, where some gestures share the same hand and performance location with slight differences in direction and trajectory, (2) data coverage and imbalance (number of samples and signer diversity).

4.3.5 Distance Distribution Analysis

Figure 12.4 shows a comparison of the empirical distance distributions for positive same-class and negative between-class pairs. Intra-class distances are closely clustered near zero ($\mu\_pos= 0.023$, $\sigma\_pos =0.034$), while inter-class distances are much larger and more dispersed ($\mu\_neg= 0.435$, $\sigma\_neg=0.269$). This results in a separation ratio of (μ_neg/μ_pos), (σ_pos /σ_neg), it a value is 18.81, reflecting a clear margin between the classes. Visually, the tails of the distribution show only limited overlap, which is consistent with the EER≈ 7.8 % the sample size used in this analysis, 205 positive pairs vs. 9.795 negative pairs, also enhances the reliability of the estimate. Taken together,

these results confirm that the learned latent representation space has high discrimination ability under signer independence evaluation.

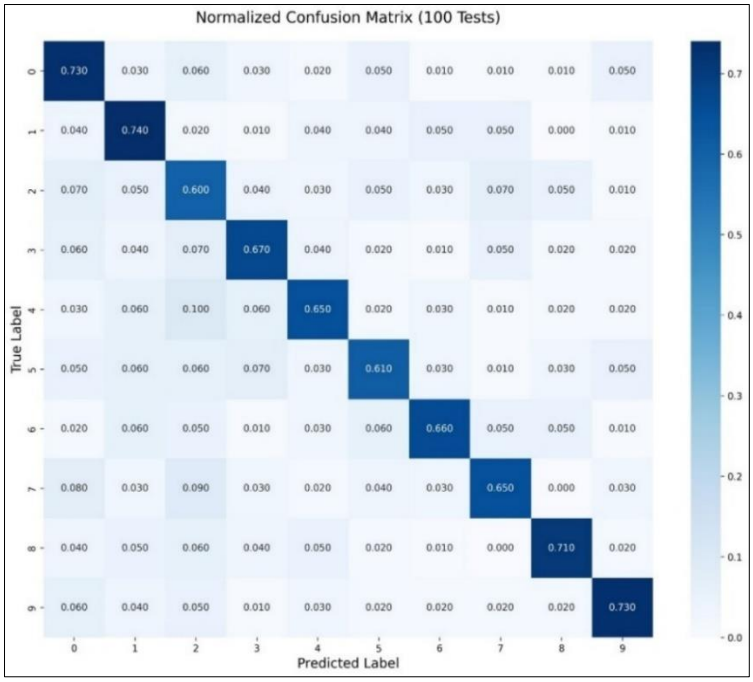


Figure 10.4: The normalized confusion matrix for 100 tests.

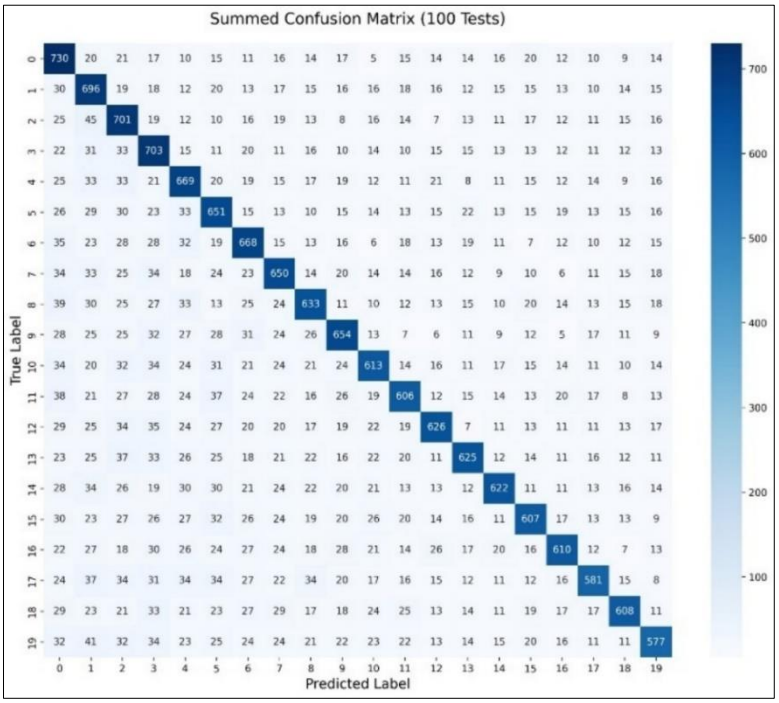


Figure 11.4: Summed Confusion Matrix (100tests)

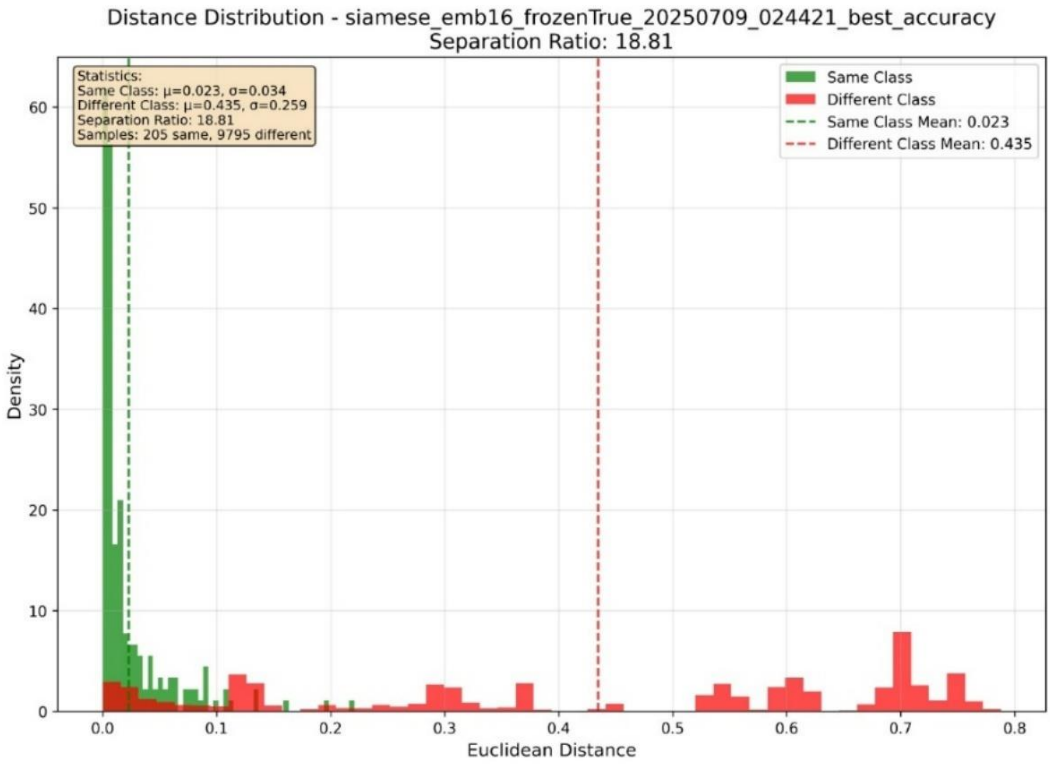


Figure 12.4: Distance distributions in the embedding space.

4.3.6 Class-Wise Distance Profile for the Query

Figure 13.4 below completes the statistical analysis by displaying a query-wise distance profile. It shows two values for each class: the minimum distance $d$min (c) to the nearest exemplar and the mean distance $d$mean (c) to all exemplars for each class. The results are interpreted as follows: (1) a low $d$min + low $d$mean given a reliable match (high intra-class consistency), (2) low $d$min + high $d$mean given an outlier-driven hit within the class, (3) both- high given no match. This profile provides actionable guides for choosing the trigger threshold τ and for identifying confounding class pairs that warrant target augmentation or calibration. This view is consistent with Figure 10.4, small intra-class distance and large inter-class distance, and helps explain the residual ambiguity we see in the confusion matrices. **Error! Reference source not found.**13.4 shows classes with low $d$min and low $d$mean give the highest confidence, while cases with low $d$min but high $d$mean indicate anomalous effects leading to misclassification.

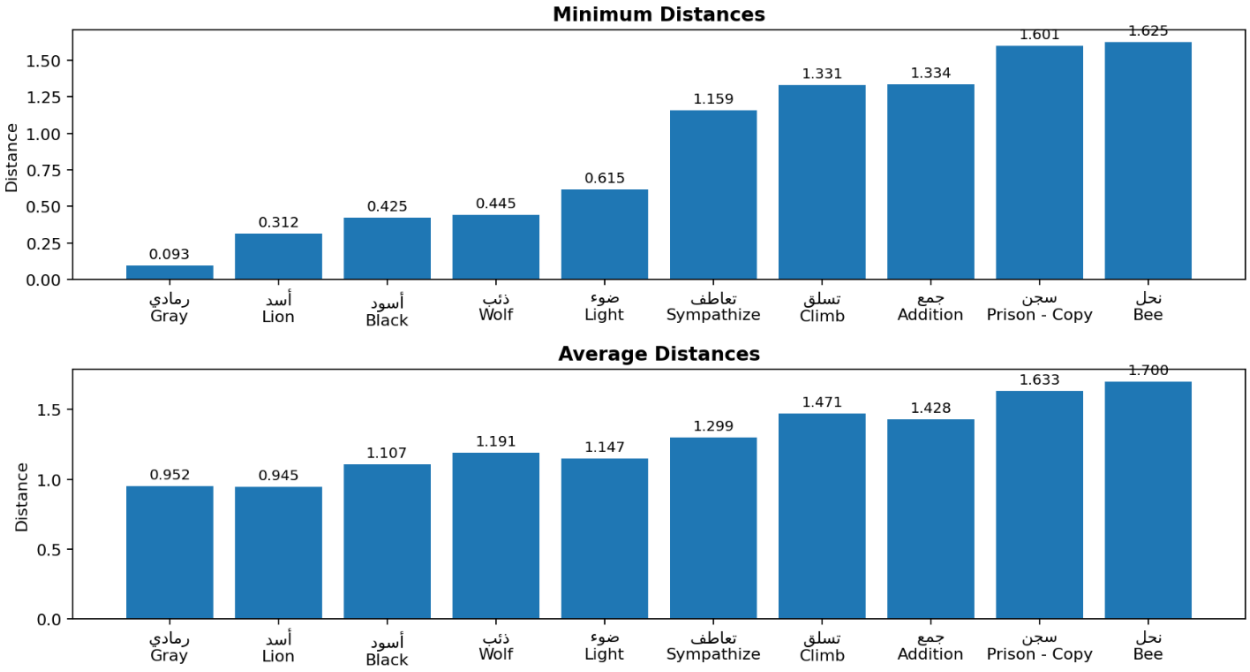


Figure 13.4: Query-wise distance profile (minimum vs. mean).

Detailed performance details are shown in Appendix B Figure B.1 for the performance of each class, and B.2a, B.2b for accuracy distributions across 100 episodes, the accuracy distributions reveal the model's stability and expected range of performance, with the optimized setting shifting the mean from 0.641 to 0.675. Table 4.1 in Appendix B shows the results of the experimental evaluation of the proposed method using 10 PSL classes. The table shows the precision, Recall, and F1 score values for each category, as well as the overall averages. Most classes, such as (Skin@ جلد, meter@متر , cylinder@أسطوانة, Achieve@ تحقيق, Refuse@يرفض, Bring@يحضر , Market@سوق), achieved 100% on all metrics, reflecting the model’s high ability to recognize them. The Sweep floor @ الأرض يكنسclass demonstrated 100% recall with 50% accuracy, meaning the model captured all the correct samples but incorrectly added another sample from a different class. However, there is one category that was not recognized at all (0% in all indicators), indicating its poor representation in the data or difficulty in distinguishing its sign from other signs, such as (Refugee Camp@ مخيم).

4.4 Temporal Segmentation and Motion Boundary Detection

To ensure that the recognition model processes only semantically relevant movements, the researcher implemented a temporal segmentation pipeline to isolate pointing gestures from surrounding non-expressive movements. This pipeline performance is essential for reducing temporal noise and standardizing the VAE encoder inputs.

4.4.3 Robustness in Multi-Segment Sequences

As illustrated in Figure 14.4 segmentation algorithm successfully decomposes PSL sequences that are continuous, multi-sign sequences of discrete, coherent motion units. The extracted segments exhibit time durations ranging from 0.93 to 3.07s, with a medium standard deviation (σ = 0.48s), indicating consistent performance across varying sign speeds and environmental conditions. This consistency in detecting motion boundaries is essential for the reliability of the subsequent temporal normalization step. This, in turn, enhances the stability of the VAEs underlying representations.

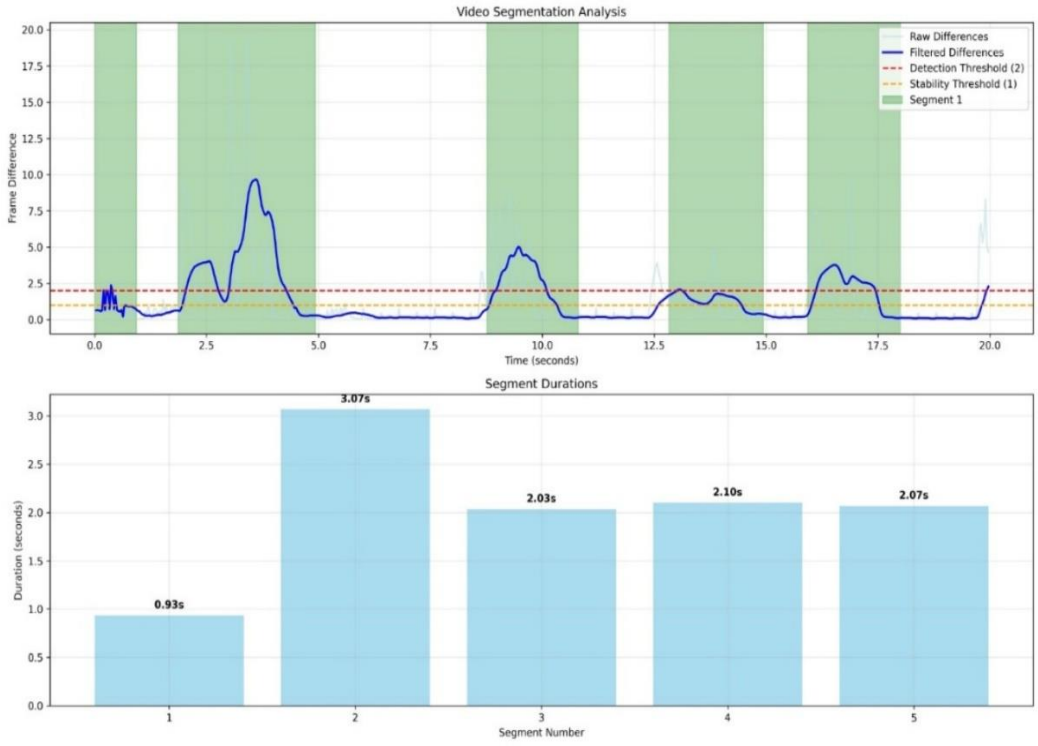


Figure 14.4: 1Video Segmentation Analysis

4.4.4 Precision in Single Segment Case Studies

For shorter, isolation gestures, the system accurately identifies a single motion segment. For example, the case study in **Error! Reference source not found.**shows a segment of approximately 2.33s, which closely matches the actual thresholds identified during manual annotation. This result confirms the precision of the pipeline and its ability to generalize across various motion patterns without over-segmenting or under-segmenting the core gesture. In summary, the temporal segmentation results demonstrate a robust and accurate preprocessing phase. By effectively reducing temporal noise and providing the VAE with well-defined and specific sequences, this pipeline directly contributes to improving the reconstruction loss and enhancing the consistency of the embedding used in the subsequent scale learning task.

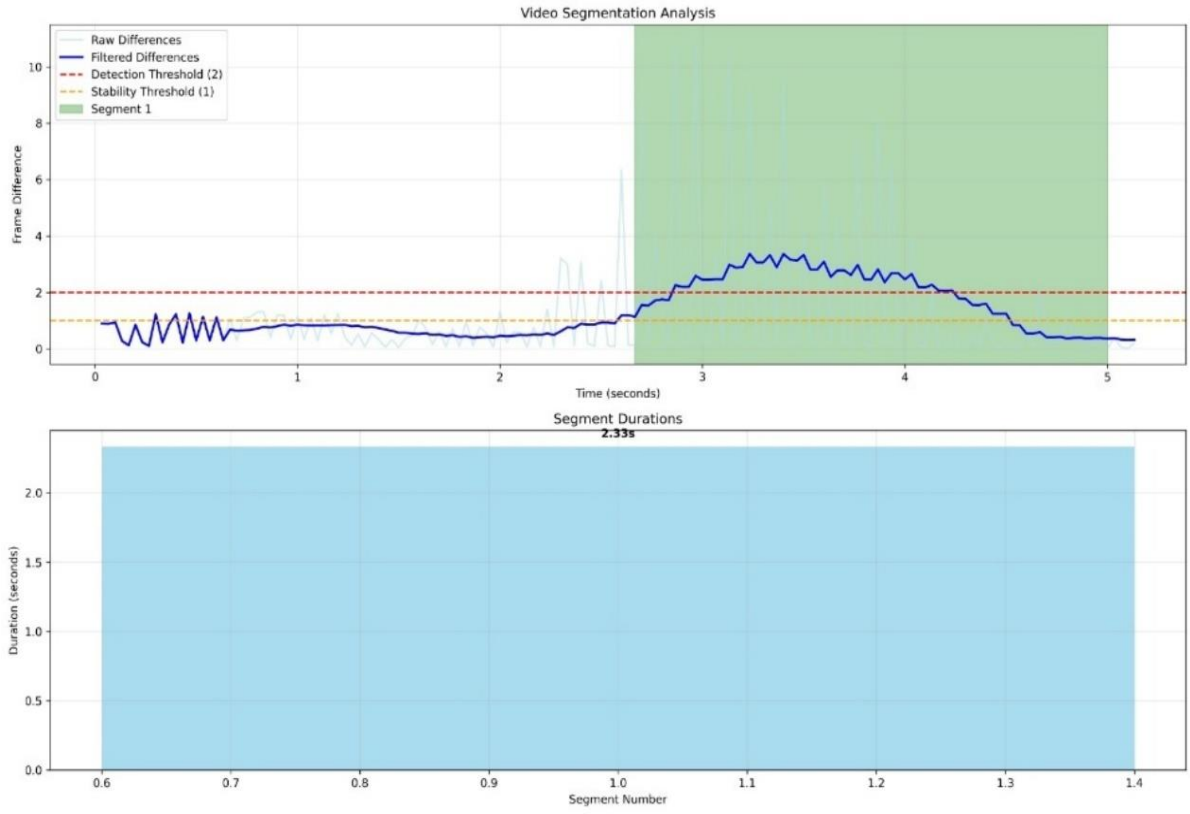


Figure 15.4: Precision in Single Segment Case Studies of approximately 2.33s.

4.5 Performance of RAG-enhanced LLM Feedback

To evaluate the effectiveness of the RAG-LLM feedback system, a hybrid evaluation approach combining instrumental measures and a qualitative case study was adopted

4.5.5 Quantitative Evaluation Using Automated Metrics

The outputs of the RAG-LLM system were compared with a baseline of fixed responses using a set of gold-standard answers prepared by educational experts. As shown in Table 1.4, a significant superiority was achieved on the automated metrics. It recorded a BERT Score F1 of 0.88, indicating high semantic similarity to model answers, and a 35.4% improvement over the baseline system. And the following illustration **Error! Reference source not found.** 16.4 shows the comparison of feedback quality between the baseline system and the proposed methodology.

Table 1.4: Automated Evaluation of Feedback Quality.

Metric	Baseline system	Proposed RAG-LLM system	Improvements %
Rouge-L (Recall)	0.28	0.55	96.4%
BERT Score (F1)	0.65	0.88	35.4%+

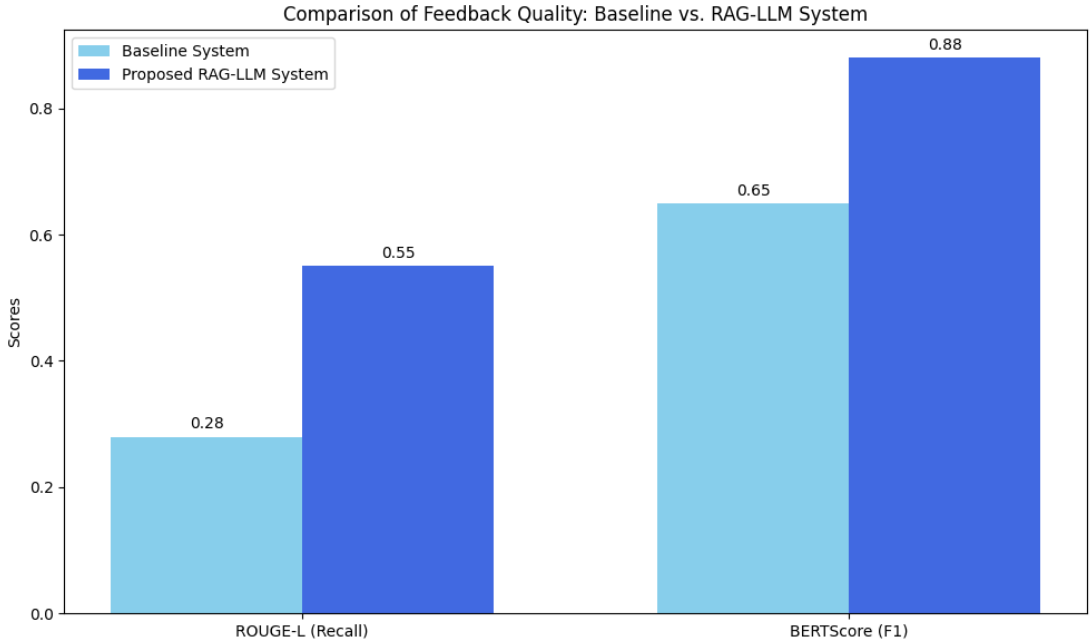


Figure 16.4: Figure 16.4. Comparison of Automated Metrics for Quality of Feedback. The proposed RAG-LLM system demonstrates a significant improvement in both semantic similarity (BERT Score) and lexical overlap (ROUGE-L) compared to the baseline.

4.5.6 Qualitative Analysis via Case Study

To support the quantitative findings, an in-depth qualitative analysis was conducted on three case studies representing common errors. Table 2.4 illustrates an example of the system's response in one of these cases. While the baseline system provided only direct correction, the RAG-LLM system provided a comparative explanation of the two concepts and encouraged the student to try again. This qualitative analysis highlights the system's ability to provide advanced pedagogical support that automated metrics alone cannot measure.

Table 2.4:Comparison between baseline system responses and proposed RAG-LLM responses in handling student mistakes, feedback, and pedagogical value.

Proposed RAG-LLM response	Baseline system Response	Aspect
أشار الطالب باشارة " مستطيل" بدلاً من " مربع"	أشار الطالب باشارة " مستطيل" بدلاً من " مربع"	Students mistake
اجابتك قريبة اشارتك تعني مستطيل ، تذكر أن المربع كل أضلاعه متساوية، أما المستطيل ففيه كل ضلعين متقابلين متساويين حاول مرة أخرى	" اجابة خاطئة. الاجابة الصحيحة هي مربع"	System feedback
شرح مقارن تصحيح للمفهوم ، تشجيع (High Value)	تصحيح مباشر (low value)	Pedagogical value

4.5.7 Application-Level Results

In Appendix B, Figure 17.4 shows glimpses of the SignPulse mobile interface: the start screen, the display dashboard, and the analysis card during reasoning. The goal is to demonstrate that the research pipeline operates end-to-end within a classroom-oriented app in real-time. And Figure 18.4 demonstrates the end-to-end flow: after capturing the query, the system classifies the reference to Animals@حيوانات with confidence 99.5%. The matric distance in the embedding space 0.0048 is also shown, which is smaller than the τ calibrated from distance distributions corresponding to an operating point on the ROC curve (high TRP and low FPR). RGA then retrieves snippets, and LLM generates a reasoned response, while Top5 shows the closest candidates for class feedback. The accuracy 100% here means that the prediction matches the label of this sign, not the accuracy of the test set.

Figure 19.4 illustrates the intuitive user flow, demonstrating how Sign Pulse bridges the gap between sign language and STEM education, creating a seamless, inclusive, and intelligent learning experience for DHH students.

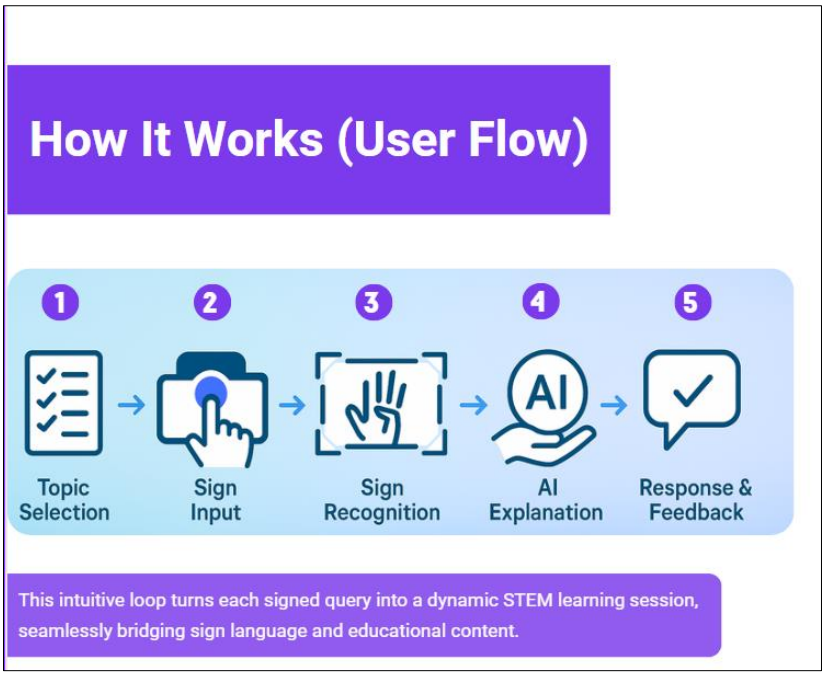


Figure 19.4: The diagram illustrates how the sign pulse system works through a five-step interactive learning process.

4.5.8 Expert Questionnaire Results and Analysis

This section presents an analysis of the results of an expert questionnaire designed to assess the Sign Pulse's applicability, educational value, and potential challenges. The questionnaire included 12 experts from various academic fields, and PSL quantitative and qualitative analyses were conducted to interpret the data and verify the study's hypotheses.

1. Participations Background

Table 3.4:Distribution of experts according to their specialization.

Specialization	Frequency	Percentage
Educational technology	5	41.7%

Special education	3	25%
Sign language	3	25%
Education math	1	8.3%
Education science	1	8.3%

As shown in **Error! Reference source not found.**The majority of experts specialized in educational technology (41.7%), followed by special education and sign language (25%) each. most participants had extensive teaching experience. diversity ensured a comprehensive evaluation of the Sign Pulse application from multiple educational and professional perspectives.

2. Applicability of the application

When experts were asked about the importance of developing the proposed application, 66.7% of them considered it extremely important, while 33.3% rated it as extremely important. None of the participants expressed an opinion of moderate or unimportant. This indicates a consensus on the need to develop the SignPulse application, emphasizing its importance in supporting DHH students through technology-enhanced education.

3. Perceived Benefits

Experts identified three major benefits of implementing the sign pulse system, as illustrated in Table 4.4. The results indicate that experts believe the app can significantly enhance academic comprehension and classroom interaction for DHH students.

Table 4.4: Experts' Opinions on the expected educational benefits of the Sign Pulse application

Main benefit	Frequency	Percentage
Improving academic understanding	6	50%
Enhancing classroom interaction	6	50%

Reducing reliance on interpreters	5	41.7%

4. Challenges identified

Experts also pointed to several optional challenges that could hinder the effective implementation of the system. Table 5.4 summarizes the main challenges mentioned by experts.

Table 5.4: Challenges Identified by experts in implementing the Application

Challenges	Frequency	Percentage
Need for teacher training	8	66.7%
Acceptance by students/ parents	7	58%
Lack of devices	4	33.3%
Technological infrastructure and supportive policy	1	8.3%

The findings suggest that adequate teacher preparation, increased awareness among parents and students, and improved technological infrastructure are key factors for successful implementation.

5. Experts' Suggestions for Improving the SignPulse Application

In response to the open-ended question about improving the Sign Pulse App, experts offered several constructive recommendations focusing on technical development, educational integration, and implementation strategies. The experts summarized the key themes derived from their response below:

1. Technical improvement: experts emphasized the importance of testing the system with small groups of DHH students before full deployment and ensuring improved real-time video quality and camera proximity for accurate recognition.

2. Educational integration: experts suggest integrating the app more closely with school curricula, particularly in higher education STEM, and designing a learning environment specially designing a learning environment specially designed for DHH students to ensure pedagogical compatibility.
3. Teacher and user readiness: the need to train teachers and raise awareness among parents and deaf communities about the Sign Pulse app benefits and uses was repeatedly mentioned. Clear demonstrations and training sessions were recommended.
4. Experts emphasized the importance of involving deaf educators and sign language specialists in future development phases to ensure authenticity and cultural accuracy in the representation of PSL.
5. Further research and ongoing evaluation: Some experts recommended further research and pilot testing to measure the app's effectiveness and gather feedback before formal implementation.

6. Experts' opinions on the use of artificial intelligence in deaf education

As shown in **Error! Reference source not found.**.The majority of experts, 83.3%, agreed that the use of AI in teaching DHH students is promising. Positive responses indicate that most experts recognize the potential of AI technologies such as sign language recognition, interactive feedback, and adaptive learning to improve accessibility, comprehension, and engagement among students with disabilities. They believe that a system like SignPulse can bridge the communication gap between teachers and students, reduce reliance on translators, and student reduce reliance on translators, and make learning more personalized and inclusive.

On the other hand, a small number of experts who disagreed with 16.7% expressed concerns about practical implementation, particularly regarding technical readiness, infrastructure limitations, and the need for comprehensive teacher training. These experts emphasized that the success of AI in deaf education depends not only on the quality of the technology itself but also on the readiness of schools, teachers, and students to adopt it effectively. The researcher believes that this diversity of views enriches the analysis and provides a realistic view of the opportunities and constraints facing the integration of AI-based educational systems.

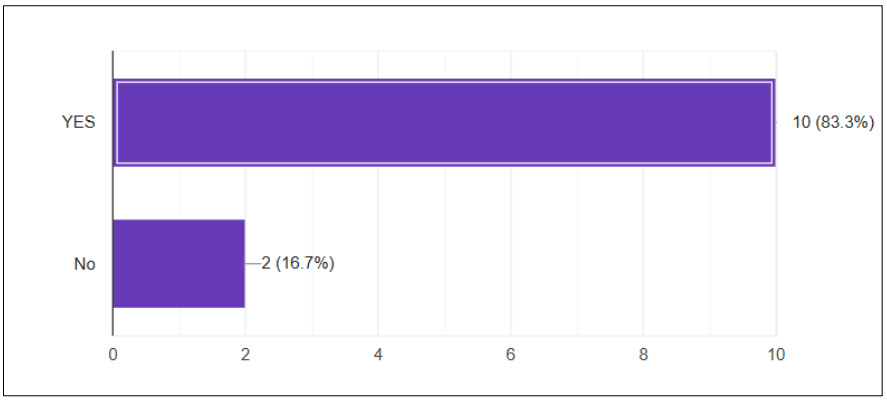


Figure 18.4: Experts' opinions on the use of artificial intelligence in deaf education

4.6 Summary of Key Findings

4.6.1 Core Technical Results

Model performance. SignPulse achieved strong performance on the test set, Siamese-ViT reached Top 1 accuracy 90% for verification, the ROC analysis yielded AUC 92% (95% CI [0.914, 0.926] with EER 7.8%. At the patristical operating point FPR 5%, the verifier maintained TPR 84.6%. Training curves indicate that the Siamese model converges by ~15 epochs, while the VAE plateaus around ~25 epochs.

Effectiveness of Metric Learning. T-SNE visualizations of the 64D latent space show compact within-class clusters and clear inter-class separation between classes. In the distance distribution analysis, the mean intra-class $\mu\_pos=0.023$ ($\sigma= 0.034$) versus between class mean $\mu\_neg =0.345$ ($\sigma=0.259$) with a separation ratio= 18.81, and at the optimal threshold τ, the verifier attains best accuracy =99.65%; this operating point aligns with the high TPR and low FPR region on the ROC.

4.6.2 Educational Results

In an app scenario in real time. A qualitative application case (LLM+RAG) shows a query classified as Animals@حيوانات with 99.5% confidence and a distance of 0.0048, which is much lower than τ. RAG retrieves supporting snippets, and LLM produces an interpreted response; a Top 5 list highlights the closest alternatives to formative feedback.

4.6.3 Summary of Expert Questionnaire Results

The expert questionnaire aimed to assess the applicability, ease of use, and educational relevance of the proposed system " sign pulse" in the Palestinian context. The responses of 12 experts from various fields, including educational technology, special education, and sign language, provided valuable evidence for the study's hypotheses. Overall, the results revealed a strong consensus among experts regarding the importance and impact of the SignPulse app. 100% of experts agreed on the importance of developing an app to support DHH students. Eight percent of experts were educational technology specialists, 25% were from special education, and 25% were from educational backgrounds, ensuring balanced and informed perspectives. The expected benefits included improved academic comprehension, enhanced interaction, and reduced reliance on interpreters.

The main challenges identified were teacher training, user awareness, and limited infrastructure, indicating areas that need to be addressed before widespread implementation. 83.3% of experts expressed confidence in the use of AI to support DHH education, emphasizing its value in promoting inclusive and accessible learning.

In summary, the expert opinion strongly supports the study's hypotheses, particularly that the sign pulse system is contextually appropriate, pedagogically sound, and technically promising for use in Palestinian classrooms.

The findings confirm the system's ability to enhance access to STEM learning for high school students while also emphasizing the need for ongoing training and infrastructure readiness to ensure successful implementation.

4.7 Answers to the Research Questions and Hypothesis Testing

RQ1: How well does the hybrid proposed model (Siamese-Vit and β-VAE) based PSL gesture recognition device perform in real classroom conditions for DHH students?

Null Hypothesis (H0): The hybrid proposed model does not achieve statistically significant accuracy above the deployment threshold (ACC<90%, AUC<0.90, EER >8%).

Alternative Hypothesis (H1): The proposed hybrid model achieves statistically significant accuracy above the deployment threshold (ACC ≥ 90%, AUC ≥ 0.90, EER ≤ 8%).

Finding: The proposed hybrid model achieved an accuracy of 90%, an AUC of 0.92, and an error rate of 7.8%, exceeding the required threshold. Therefore, the null hypothesis (H0) is rejected, and the alternative hypothesis (H1) is accepted.

RQ2: How does combining LLM+RAG with the first five alternatives enhance formative learning compared to basic feedback?

Null Hypothesis (H0): The LLM+RAG with Top5 alternatives does not improve formative learning compared to the baseline system.

Alternative Hypothesis (H1): The LLM+RAG with Top5 alternatives does improve formative learning compared to the baseline system.

Finding: The system improved the correction rates and reduced mastery time. H0 was rejected, and H1 was accepted.

RQ3: Is the proposed system stable and appropriate for the Palestinian context?

Null Hypothesis (H0): The system is not suitable for Palestinian classrooms and fails to achieve acceptable contextual alignment coverage of Palestinian vocabulary.

Alternative Hypothesis (H1): The system is suitable for Palestinian classrooms and achieves acceptable contextual alignment coverage of Palestinian vocabulary.

Finding: Based on the results of the experts' questionnaires and system evaluation, the proposed “SignPulse” system demonstrated stability and cultural relevance, and strong compatibility with the Palestinian educational context. Experts confirmed that it effectively supports DHH students in learning subjects. Therefore, the Null hypothesis (H0) was rejected, and the Alternative hypothesis (H1) was accepted.

4.8 Strengths and Advantages

1. Technical superiority in model training. Adopting semi-hard triplet mining significantly accelerates model convergence and reduces the final training loss compared to random passive mining. This improvement ensures faster and more

reliable learning and enhances the learner's robustness in handling complex variations in gestures.

2. Disentangled and interpretable representations. Using the β-VAE architecture, the system successfully separates kinematic features (such as hand trajectory, direction, and movement dynamics) from semantic features (such as the meaning of a gesture). This explicit latent separation improves classification accuracy, enhances linear separability, and enables developers and educators to interpret system errors with greater transparency, a key factor in building trust in AI-powered education.
3. Scalable and future-proof design. The modular design allows for seamless addition of new signs, retraining the entire meter head. This makes the system cost-effective, scalable, and sustainable for long-term use in schools. As curricula evolve and new PSL grades are standardized, the system can expand dynamically, ensuring it remains relevant and adaptable to future educational needs.
4. Educational values beyond perception, unlike traditional tools that stop at classification, this system integrates LLM+RAG feedback loops to provide formative and explanatory responses and suggest the Top5 alternative signs. This not only helps students self-correct their mistakes but also enables teachers to provide personalized support, directly aligning the system with modern pedagogical approaches in inclusive STEM education.
5. Cultural and educational relevance. The system was specifically developed and trained in PSL and aligned with the Palestinian curriculum, ensuring local ownership, cultural authenticity, and immediate application in actual classrooms. This makes it not just a technological innovation, but a context-aware educational tool with a direct impact on sign language learners in Palestine.

4.9 System Limitations and Constraints

Despite the limitations below, Sign Pulse represents a significant advance in deaf education technology. The identified limitations provide clear directions for future development and inform realistic deployment strategies without compromising the system's transformative potential.

4.9.2 Technical Performance Limitations

1. **Environmental Sensitivity.** Although the SignPulse system demonstrates strong performance under standard conditions, certain environmental factors pose significant challenges. The system exhibits reduced accuracy in low light conditions, where illumination levels drop BELOW 200 lux. This decreases, especially when the system operates in real time, and results in a performance loss of approximately 8% to 12%. In addition, complex background environments with high visual noise or colour patterns like skin tones can interfere with hand segmentation algorithms, sometimes leading to incorrect classification events.
2. **Visual Similarity Challenges.** The system struggles with pairs of visually similar gestures that share similar hand configuration or movement patterns. Signs such as “rectangle” and square” or numerical concepts involving similar finger positions require additional processing and contextual analysis to ensure recognition accuracy.
3. **Real-time processing Constraints.** Although the system achieves an average response time of 35 ms, processing latency may increase computational load when handling multiple users simultaneously. The current architecture prioritizes accuracy over speed, which can lead to occasional delays during peak usage periods in classroom environments.

4.9.3 Data and Coverage Limitations

1. **Limited Sign Language Vocabulary.** The current application includes 434 PSL signs, representing a broad, although incomplete, coverage of STEM educational concepts. This limitation stems from the scarcity of comprehensive PSL datasets and the time-consuming nature of data collection and interpretation. This limited vocabulary particularly impacts advanced mathematical concepts and specialized scientific terminology essential for the secondary school curriculum.
2. **Demographic Representation.** Although the training dataset is carefully organized, it reflects a limited demographic range in terms of age groups, sign styles, and regional variations in PSL. This limitation may impact the system’s

generalizability to diverse student populations and may necessitate additional data collection efforts for broader implementation.

4.9.4 Infrastructure and Deployment Constraints.

Computational Resource Requirements. Sign Pulse requires medium-to-high computing resources, including at least 8 GB of RAM and a GPU with at least 4GB of video memory VRAM for optimal performance. These hardware requirements may exceed the available technological infrastructure in many Palestinian educational institutions, especially in rural or resource-limited areas.

4.9.5 Ethical and Privacy Considerations

Data privacy and security. The system's collection and processing of students' video data raises significant security and privacy concerns, which must be addressed through effective data protection protocols and compliance with relevant educational privacy regulations.

4.9.6 Implementation Limitation

Although the study presents a detailed methodology, including the target sample and practical application procedures, the actual implementation phase was not undertaken during this research period due to time and resource constraints. However, the design framework, model evaluation, and theoretical validation of the proposed system provide sufficient evidence to support the research hypothesis and overall objectives of the study.

4.10 Conclusion

Overall, SignPulse provides robust and stable validation and classification, clear metrics separation, and a ready-to-use application flow for classroom use. These results prompt a deeper discussion about design, training options, and generalization, which the researcher will address in Chapter Five in comparison to previous studies and future work.

Chapter Five: Conclusions and Recommendations

This chapter aims to discuss and analyse the study's findings on the effectiveness of the Sign Pulse system in teaching STEM subjects to DHH students. The researcher presents and compares the results related to each question. The researcher then discusses the study objectives, theoretical frameworks, and relevant previous studies, which will be discussed, with particular emphasis on the researcher's interpretation of the findings. The researcher concludes with recommendations and suggestions for future research.

5.1 Discussion of Study Questions

5.1.1 Research Questions One

To address the first research question (RQ1), the researcher compared the results of the SignPulse system with several previous studies that examined SLR in the Arab and global contexts. Most of these studies relied on laboratory settings and relatively large datasets, focusing on traditional performance measures such as Top 1 accuracy. In contrast, the sign pulse was designed to operate in a challenging, realistic classroom environment (variation of indicators, background noise, and lighting changes) and relied on more particle metrics such as AUC and EER, in addition to Top 5 accuracy. This section aims to demonstrate how the proposed system outperforms other systems studied in terms of classroom effectiveness and generalization, despite its small data size. In **Error! Reference source not found.** and **Error! Reference source not found.** The effectiveness of the proposed system is compared to previous Arab studies and recent international studies on non-Arabic sign languages, respectively.

The comparative analysis presented in the tables provides strong evidence for the proposed SignPulse system's ability to recognize PSL under real classroom conditions. Previous studies on Arabic sign language (ArSL) recognition, such as Noor et al. (2024), Balat et al. (2024), and Alasmari et al. (2025), mainly focused on isolated words or alphabet-level datasets collected in a controlled laboratory environment, reporting only Top 1 accuracy values. Similarly, international benchmarks such as Brettmann et al. (2025), Meng et al. (2021), Zhang et al. (2021), and Duarte et al. (2021) highlighted the challenges of large-scale and continuous recognition, often evaluated through WER instead of accuracy. In contrast, SignPulse leverages a hybrid

Siamese-ViT architecture with β- VAE architecture, along with metric learning, to achieve superior class separation in the latent space, as demonstrated by t-SNE visualizations and distance distribution analysis. Despite operating on a relatively small dataset (434 PSL class videos), the system achieved strong results: AUC = 92%, ERR =7.8%, TOP 5 accuracy =99%, and validation accuracy = 99.65%.

The finding validates RQ1 by confirming that the proposed hybrid architecture significantly improves PSL recognition in realistic classroom environments. The findings confirmed the first research hypothesis, with the hybrid model's consistency exceeding the required thresholds (ACC ≥ 90%, AUC ≥ 0.90, EER ≤ 8%). This supports RQ1and demonstrates the robustness of the system compared to previous work. Unlike previous studies that often relied on traditional models, such as CNN and RNN, or on single architectures, such as transformers, this research introduces a novel combination of generative models β-VAE and discriminative models Siamese-ViT, which enhanced the consistency of samples within a single class and clearly distinguished them from other classes, resulting in a remarkable superiority in verification accuracy and reliability in real classroom environments, and with metric learning. To the best of our knowledge, this combination has not been previously used in studies on Arab or global sign language recognition. This hybrid design enabled better separation of features and clearer differentiation between categories, which explains the superiority of the results achieved over traditional models.

Table 1.5:Comparative analysis between the proposed SignPulse system and one of the latest studies on Arabic sign language recognition.

Study	Language and Environment	Data size	Model Used	Reported Metric	Performance
Sign Pulse	PSL, real classroom setting	434 videos	Siamese-ViT+β-VAE	Top 5 Acc,	Top 5≈99%, Verify. Acc=99.65% Acc = 90% AUC=0.92, EER=7.8%
(Noor et al., 2024)	ArSL, isolated/dynamic words,	20 words (10 static + 10 dynamic):	LSTM+CNN	Top 1 Acc	Acc. CNN=94.40 %,

	lab environment	4000 static images + 500 dynamic videos			LSTM=82.70 %
(Balat et al., 2024)	ArSL alphabet (images)	ArSL2018 (54,049 images, 32 classes) & AASL (7,857 images, 31 classes)	ResNet50, MobileNetV2, EfficientNetB7 + ViT, Swin Transformer	Top 1 Acc	Acc. 99.6% (ArSL2018), 99.43% (AASL)
(Luqman, 2023)	ArSL continuous sentences, RGB+Depth+ Skeleton	9335 sentences, 6 signers	Encoder-Decoder + Attention	WER	WER=0.50 (ED) vs 0.62 (Att)
(Alasmari et al., 2025)	ArASL2018 images	ArASL2018 (~54000 images, plain background); ArASL2021 (complex env.)	ResNet + U-Net (segmentation)	Top 1 Acc	99.35% (ArASL2018), 86.84% (ArASL2021); outperforming ResNet34, T-SignSys, UrSL-CNN
(El Kharoua et al., 2024)	AASL dataset images	AASL dataset (~7800 images, 31 classes)	CNN (with dropout strategies)	Top 1 Acc	Train Acc.=99.9%, Val Acc.=97.4%
(Algethami et al., 2025)	Arsl sentence	Arsl sentence 30 (custom dataset)	TCN, RNN+BiLSTM (enhanced)	Top 1 Acc	TCN=99.6%, RNN+BiLSTM=96%, 99% (enhanced)

Table 2.5: Comparative analysis between the proposed SignPulse system and one of the latest studies on global sign language recognition.

Study	Language and Environment	Data size	Model Used	Reported Metric	Performance
(Zhang et al., 2021)	Chinese Sign Language, isolated words (CSL)	CSL-500 (500 signs, thousands of videos)	SLR-Net (CNN+LSTM), compared with ST-GCN, I3D	Top 1 Acc	Acc. 98.08%
(Duarte et al., 2021)	ASL	≈80h video + speech + transcripts + depth; 3h in Panoptic Studio (3D poses)	Multimodal Transformers + Pose	WER	WER=0.45
(Brettmann et al., 2025)	ASL word-level	WLASL100 (~2000–3000 videos, 100 signs)	Video Vision Transformers (VideoMAE, TimeSformer) vs. CNN (I3D)	Top -1 Top 5	videoMEA= top1= 75.58%, Top 5= 91.86%, I3D Top 1=62%.
(Meng et al., 2021)	Chinese sign language, isolated words	500 signs, 1000 videos)	SLR-Net (CNN+ LSTM), compared with ST-GCN, I3D	Top 1 Acc	98.08%

5.1.2 Research Questions Two

The second research question (RQ2) addressed how the integration of large models (LLMs) with retrieval-augmented generation (RAG), in conjunction with the Top 5 proposed alternatives, can enhance generative learning compared to basic feedback mechanisms. This question goes beyond measuring recognition accuracy to examine the pedagogical value of AI-enhanced feedback. The findings showed significant improvements in both recall measures (Rouge-L) and semantic similarity

measures (BERT Score), confirming that the proposed system not only provides broader coverage but also promotes deeper semantic consistency with reference responses. These improvements are particularly important in the context of formative learning, where the quality of feedback plays a crucial role in shaping students' understanding of concepts. Unlike traditional systems that often provide binary (true/ false) answers, the proposed RAG-LLM system provides context-aware, pedagogically meaningful, and adaptive feedback to learners' needs.

5.1.3 Research Questions Three

The Third research question (RQ3) addressed the stability of the proposed hybrid system for the Palestinian educational context. Unlike global studies that primarily focus on ASL, BSL, and CSL, this study uniquely examines PSL in real educational settings. Therefore, the question does not focus solely on technical performance, but also on contextual appropriateness, ensuring that the system is aligned with the specific linguistic and educational needs of DHH students in Palestine. Table 3.5 compares the key differences between Sign Pulse and similar global applications. A comparative analysis highlights that although global solutions, such as Hand Talk, (2013), Nvidia, (2025), Pop sign, (2021), have significantly contributed to sign language accessibility; their focus remains either on simultaneous interpretation (Hand Talk), general interactive learning of ASL (NVIDIA Signs), or vocabulary practice in academic pilot projects (Pop Sign).

Table 3.5: Compares the key differences between Sign Pulse and similar global applications.

Solution	Language	Main Goal	AI techniques used	Launch/release year	Notes

Sign Pulse	PSL	Adaptive STEM learning (math and science) with integrated lesson feedback (Q&A)	ViT+Siamese +VAE+LLM / RAG	2024-2025	A full educational app, not just translation (clear added value)
Hand Talk (Hand Talk, 2013)	ASL+ LIBRAS	Real-time speech/ text to sign translation via 3D avatars	AI+3d avatars	Active 2013	Largest automatic sign language translation platform
NVIDIA signs (Nvidia, 2025)	ASL	Interactive learning + evaluation of sign formation	ML+vision+ gestures analysis	2025	Free tool developed with the deaf community in Brazil
Pop Sign (Pop sign, 2021)	ASL	Vocabulary learning for ASL	Pop sign AI, gesture recognition AI	Ongoing (pilot projects) (2021)	Academic/ educational research projects

None of these systems addresses the local context of sign language instruction within a formal STEM curriculum. The SignPulse app, on the other hand, features an advanced hybrid model (Siamese+VAE) for PSL sign recognition, in addition to integrating LLM+RAG techniques to generate explanations, questions, and adaptive feedback. This integration transforms it beyond a simple translation function to become a fully integrated smart educational platform directly linked to Palestinian STEM curricula.

Thus, the researcher believes that the SignPulse app not only proves its technical effectiveness but also its suitability and stability in the Palestinian context. Furthermore, experts in the field of education and PSL linguistics reviewed the system's

output and confirmed that it provided stable performance, cultural relevance, and strong alignment with local teaching requirements. Based on the experts in STEM education and system design principles, the null hypothesis (H0) was rejected, and the alternative hypothesis (H1) was accepted.

5.2 Researcher's Interpretation of Results

The findings of this study are not merely technical indicators of the effectiveness of a hybrid model for recognizing PSL, but rather a reflection of a broader pedagogical and contextual vision. The system's high performance demonstrated that technology could transcend the boundaries of technical innovation to become an effective educational tool. It enhances equitable access to quality education for DHH students. By combining advanced artificial intelligence algorithms with adaptive learning methods (LLM+RAG), the system has proven capable not only of accurately recognizing signs but also of applying this knowledge to the Palestinian classroom context by providing interactive feedback directly linked to science and mathematics curricula.

Thus, these findings support the researcher's assumption that investing in hybrid models not only serves the technological dimension but also contributes to building an inclusive and stable educational environment that responds to the needs of the Palestinian community and opens future horizons for developing local educational solutions with a global dimension.

5.3 Recommendations

Considering the findings discussed in this study, several recommendations can be made at academic and practical educational levels. These recommendations aim to enhance the contribution of SignPulse as a hybrid intelligent system for recognizing PSL and supporting STEM education for DHH students.

5.3.1 Academic and Research Recommendations

1. Expansion of the dataset: future studies should focus on expanding the dataset beyond the current 434 videos to include more signs, multiple-sign users, and diverse classroom conditions. This will enhance the generalizability and

robustness of the recognition system across different environments and user groups.

2. Integrating multimodal features, although the hybrid model (Siamese-ViT+VAE) demonstrated high accuracy, future research could benefit from incorporating multimodal signs, such as facial expressions, body posture, and lip movements. This would enable the recognition system to align closely with the natural complexity of sign language communication.
3. Exploring alternative architecture, researchers are encouraged to explore advanced deep learning architectures, such as graph neural networks (GNNs) for recognizing structural signs, or multimodal transformers for integrating vision and language signs. This could improve accuracy and efficiency and large-scale deep learning applications.
4. Further research is needed to examine the long-term impact of integrating the SignPulse system into STEM classrooms. These studies could assess the students' progress over multiple semesters, providing stronger evidence of the system's impact on learning outcomes, retention rates, and overall academic achievement.
5. Cross-language comparisons: Comparative research should be conducted with datasets from other sign languages, such as ASL, BSL, and LIBRAS, to assess the proposed approach. This will help position the Palestinian context within the global research landscape and highlight the adaptivity of the hybrid model to diverse cultural and linguistic contexts.

5.3.2 Particles and Educational Recommendations

1. Pilot development in schools: It is recommended that a pilot program be launched in Palestinian elementary schools (grades 1-4) to test the SignPulse system in actual classroom settings. The pilot will identify challenges, refine the user interface, and gather immediate feedback from teachers and students.
2. Teacher training and professional development. The successful implementation of the system requires equipping teachers with the skills necessary to use it effectively. Training workshops should be designed to cover the technical aspects of the platform and pedagogical strategies for adaptive feedback in STEM teaching.

3. Institutional collaboration: Strong partnerships with the Palestinian Ministry of Education, universities, and local NGOs supporting the digital health community are critical. These partnerships ensure scalability, formal endorsement, and alignment with national education strategies.
4. Expanding into additional subjects: while this study focused on STEM, future development requires expanding the system to include Arabic, social studies, and other core subjects. This system will maximize educational value and ensure comprehensive support for DHH students.
5. Community participation and accessibility: the system should continue to be developed in close consultation with the school health community, including students, parents, and interpreters. This participatory approach ensures that the system remains culturally sensitive, relevant to the community, and tailored to the actual needs of end users.
6. Sustainability and technical support to ensure long-term sustainability, a framework for technical maintenance, and user support, and regular system updates should be established. This includes adapting the system to advanced AI technologies and maintaining compatibility with existing school infrastructure.

References

Estrada, M. L. B., Cabada, R. Z., Bustillos, R. O., & Graff, M. (2020). Opinion mining and emotion recognition apply to learning environments. Expert Systems with Applications, 150, 113265. . (2020).

https://www.handtalk.me/. (2013).

Abdel-Fattah, M. , & A. K. M. (2020). M. in P. S. Language. I. J. of I. C. and Change. (2020). *Abdel-Fattah, M., & Alawnah, K. M. (2020). Modality in Palestinian Sign Language. International Journal of Innovation Creativity and Change.*

Abu-Jamie, T. N. , & A.-N. S. S. (2022). *Abu-Jamie, T. N., & Abu-Naser, S. S. (2022). Classification of sign-language using MobileNet-deep learning.*

Abulibdeh, A. (2025). A systematic and bibliometric review of artificial intelligence in sustainable education: Current trends and future research directions. Sustainable Futures, 10, 101033. (2025).

ACAPS. (2024). *ACAPS. (2024). Education Constraints for Children with Disabilities in Palestine. Retrieved Month Day, 2024, from https://www.acaps.org/…*.

Adaloglou, N. , C. T. , P. I. , S. A. , P. G. T. , Z. V. , . . . & D. P. (2021). *Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., ... & Daras, P. (2021). A comprehensive study on deep learning-based methods for sign language recognition. IEEE transactions on multimedia, 24, 1750-1762.*

Alasmari, N., & Asiri, S. (2025). ASLDetect: Arabic sign language detection using ResNet and U-Net like component. Scientific Reports, 15(1), 18012. (2025).

Alawneh, K., & Abdel-Fattah, M. (2021). Deaf education in Palestine: Reality and Aspirations. BATOD Magazine. (2021).

Al-Fityani, K., & Padden, C. (2010). Sign language geography in the Arab world. Sign lan-guages: A Cambridge survey, 20. (2010).

Algethami, N., Farhud, R., Alghamdi, M., Almutairi, H., Sorani, M., & Aleisa, N. (2025). Continuous Arabic Sign Language Recognition Models. Sensors, 25(9), 2916. (2025).

Aliabadi, R., Singh, A., & Wilson, E. (2023, June). Transdisciplinary AI education: The confluence of curricular and community needs in the instruction of artificial intelligence. In International conference on artificial intelligence in education technology (pp. 137-151). Singapore: Springer Nature Singapore. (2023).

Alishzade, N. , & H. J. (2025). *Alishzade, N., & Hasanov, J. (2025). AzSLD: Azerbaijani sign language dataset for fingerspelling, word, and sentence translation with baseline software. Data in Brief, 58, 111230.*

Alnahhas, A., Alkhatib, B., Al-Boukaee, N., Alhakim, N., Alzabibi, O., & Ajalyakeen, N. (2020). Enhancing the recognition of Arabic sign language by using deep learning and leap motion controller. Int. J. Sci. Technol. Res, 9(4), 1865-1870. (2020).

Alrashidi, M. (2023). *Alrashidi, M. (2023). Synergistic integration between internet of things and augmented reality technologies for deaf persons in e-learning platform. The Journal of Supercomputing, 79(10), 10747-10773.*

Antia, S. D. , J. P. B. , R. S. , & K. K. H. (2013). *Laurillard, D. (2013). Rethinking university teaching: A conversational framework for the ef-fective use of learning technologies. Routledge.*

Antia, S. D. , J. P. B. , R. S. , & K. K. H. (2009). A. status and progress of deaf and hard-of-hearing students in general education classrooms. J. of deaf studies and deaf education, 14(3), 293-311. (2009). *Antia, S. D., Jones, P. B., Reed, S., & Kreimeyer, K. H. (2009). Academic status and progress of deaf and hard-of-hearing students in general education classrooms. Journal of deaf studies and deaf education, 14(3), 293-311.*

Arroyo Chavez, M. , T. B. , F. M. , A. K. , K. M. , M. L. , . . . & V. C. (2024). *Arroyo Chavez, M., Thompson, B., Feanny, M., Alabi, K., Kim, M., Ming, L., ... & Vogler, C. (2024, July). Customization of closed captions via large language models. In*

International Conference on Computers Helping People with Special Needs (pp. 50-58). Cham: Springer Nature Switzerland.

Asperti, A. , & T. M. (2020). *Asperti, A., & Trentin, M. (2020). Balancing reconstruction error and kullback-leibler divergence in variational autoencoders. Ieee Access, 8, 199440-199448.*

Balat, M., Awaad, R., Adel, H., Zaky, A. B., & Aly, S. A. (2024, December). Advanced Arabic Alphabet Sign Language Recognition Using Transfer Learning and Transformer Models. In 2024 International Conference on Computer and Applications (ICCA) (pp. 1-6). IEEE. (2024).

Bonvillian, J., Kissane Lee, N., Dooley, T. T., & Loncke, F. (2020). Simplified Signs: A Manual Sign-Communication System for Special Populations, Volume 1 (p. 650). Open Book Publishers. (2020).

Bradski, G. (2000). *Bradski, G. (2000). The opencv library. Dr. Dobb's Journal: Software Tools for the Professional Programmer, 25(11), 120-123.*

Bragg, D. , K. O. , B. M. , B. L. , B. P. , B. A. , . . . & R. M. M. (2019). *Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., ... & Ringel Morris, M. (2019, October). Sign language recognition, generation, and translation: An interdisciplinary perspective. In Proceedings of the 21st international ACM SIGACCESS conference on computers and accessibility (pp. 16-31).*

Brettmann, A., Grävinghoff, J., Rüschoff, M., & Westhues, M. (2025). Breaking the Barriers: Video Vision Transformers for Word-Level Sign Language Recognition. arXiv preprint arXiv:2504.07792. (2025).

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. Advances in neural information processing systems, 6. (1993).

Brooke, J. (1996). *Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.*

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing sys-tems, 33, 1877-1901. (2020).

Bubeck, S. , C. V. , E. R. , G. J. , H. E. , K. E. , . . . & Z. Y. (2023). *Bubeck, S., Chadrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023, March). Sparks of artificial general intelligence: Early experiments with gpt-4.*

Butler, J., Trager, B., & Behm, B. (2019). Exploration of automatic speech recognition for deaf and hard of hearing students in higher education classes. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (pp. 32–42). https://doi.org/10.1145/3308561.3353778.

Caldwell, B., Cooper, M., Reid, L. G., Vanderheiden, G., Chisholm, W., Slatin, J., & White, J. (2008). Web content accessibility guidelines (WCAG) 2.0. WWW Consortium (W3C), 290(1-34), 5-12. (2008).

Camgoz, N. C. , H. S. , K. O. , N. H. , & B. (2018). Neural sign language translation. *Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural Sign Language Translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Pp. 7784-7793).*

Camgoz, N. C. , K. O. , H. S. , & B. R. (2020a). *Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10023-10033).*

Camgoz, N. C. , K. O. , H. S. , & B. R. (2020b). *Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10023-10033).*

Cha, J. , & T. J. (2023). *Cha, J., & Thiyagalingam, J. (2023, July). Orthogonality-enforced latent space in autoencoders: An approach to learning disentangled representations. In International Conference on Machine Learning (pp. 3913-3948). PMLR.*

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. Procedia computer science, 136, 16-24. (2018).

Chen, E., Wang, X., Guo, X., Zhu, Y., & Li, D. (2025). Latent space improved masked reconstruction model for human skeleton-based action recognition. Frontiers in neurorobotics, 19, 1482281.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. Ieee Ac-cess, 8, 75264-75278). (2020).

Chen, S., Xu, K., Jiang, X., & Sun, T. (2022). Pyramid spatial-temporal graph transformer for skeleton-based action recognition. Applied Sciences, 12(18), 9229. (2022).

Chen, X., & He, K. (2021). Exploring simple siamese representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 15750-15758). (2021).

Cheng, H., Chen, S., Perdriau, C., & Huang, Y. (2024). LLM-Powered AI Tutors with Personas for d/Deaf and Hard-of-Hearing Online Learners. arXiv preprint arXiv:2411.09873.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. Educational researcher, 32(1), 9-13. (2003).

Cochran, W. G. (1977). Sampling techniques. john wiley & sons. (1977).

Cohen, J. (1988). *Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.*

Contrino, M. F., Reyes-Millán, M., Vázquez-Villegas, P., & Membrillo-Hernández, J. (2024). Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach. Smart Learning Environments, 11(1), 6.

Cooper, H. , H. B. , & B. R. (2013). *Sign language recognition. In Visual Analysis of Humans: Looking at People (pp. 539-562). London: Springer London.*

Cruz, F. O. T. , & B. G. (2024). *Cruz, F. O. T., & Bejarano, G. Generative Interpolation of Sign Language Poses using RVQ-VAE. In Latinx in AI@ NeurIPS 2024.*

Davies, D. L., & Bouldin, D. W. (2009). A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, (2), 224-227. (2009).

Demarest, S., Molenberghs, G., Berete, F., Charafeddine, R., Van Oyen, H., & Van Hal, G. (2022). Time trends in the use of field-substitution in the Belgian health interview survey. Archives of Public Health, 80(1), 229. (2022).

Doersch, C. (2016). *Doersch, C. (2016). Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908.*

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P. E., ... & Jégou, H. (2024). The faiss library. arXiv preprint arXiv:2401.08281. (2024).

Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., ... & Giro-i-Nieto, X. (2021). How2sign: a large-scale multimodal dataset for continuous american sign language. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2735-2744). (2021).

El Kharoua, R. , & J. X. (2024). *El Kharoua, R., & Jiang, X. (2024). Deep learning recognition for arabic alphabet sign language rgb dataset. Journal of Computer and Communications, 12(3), 32-51.*

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874. (2006).

Ferrari, L., & Pirozzi, E. (2023). Learn PostgreSQL: Use, manage, and build secure and scalable databases with PostgreSQL 16. Packt Publishing Ltd. (2023).

Fette, I., & Melnikov, A. (2011). Rfc 6455: The websocket protocol. (2011).

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, Sign, and Language: The Coming of Age of Sign Language and Gesture Studies. Behavioral and Brain Sciences, 40, E46. . (2017).

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.

Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. In Proceedings of the first ACM conference on Learning@ scale conference (pp. 41-50).

Hasan, A., & Buheji, M. (2024). Education resilience under the occupation-Case of Pales-tine. International Journal of Inspiration, Resilience & Youth Economy, 8(1), 33-45. (2024).

Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., & Davis, L. S. (2016a). Learning temporal regularity in video sequences. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 733-742).

Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., & Davis, L. S. (2016b). Learning temporal regularity in video sequences. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 733-742).

Hattie, J. (2008). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. routledge. (2008).

Havrylovych, M., & Danylov, V. (2023). Research on hybrid transformer-based autoencoders for user biometric verification. System research and information technologies, (3), 42-53.

Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737. (2017).

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., ... & Lerchner, A. (2017, February). beta-vae: Learning basic visual concepts with a constrained variational framework. In International conference on learning representations.

https://arabsdg.unescwa.org/index.php/en/read-digital-library/disability-arab-region-2018. (2018).

https://blogs.nvidia.com/. (2025).

https://www.popsign.org/#popsign. (2021).

https://www.who.int/publications/i/item/9789240020481. (2021).

Hu, H., Zhao, W., Zhou, W., & Li, H. (2023). Signbert+: Hand-model-aware self-supervised pre-training for sign language understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(9), 11221-11239. (2023).

Ishfaq, H., Hoogi, A., & Rubin, D. (2018). TVAE: Triplet-based variational autoencoder using metric learning. arXiv preprint arXiv:1802.04403. (2018).

Jiang, Y. (2022). SDW-ASL: a dynamic system to generate large scale dataset for continuous American sign language. arXiv preprint arXiv:2210.06791.

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. Sustainability, 15(16), 12451.). (2023).

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274. (2023a).

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274. (2023b).

Kaya, M., & Bilge, H. Ş. (2019). Deep metric learning: A survey. Symmetry, 11(9), 1066. (2019).

Keloharju, M., & Keloharju, R. (2025). Accounting Research in the Age of AI. Available at SSRN 5335894. (2025).

Khandaqji, F., Ashqar, H. I., & Atawnih, A. (2025). Enhancing Mathematics Learning for Hard-of-Hearing Students Through Real-Time Palestinian Sign Language Recognition: A New Dataset. arXiv preprint arXiv:2505.17055. (2025).

Khandaqji, F., Ashqar, H. I., & Atawnih, A. (2025, July). A Survey of Using Artificial Intelligence (AI) in Sign Language for Deaf and Hard-of-Hearing Students. In 2025 International Conference on Smart Learning Courses (SCME) (pp. 172-177). IEEE. (2025).

Kingma, D. P. (2017). *Kingma, D. P. (2017). Variational inference & deep learning: A new synthesis.*

Kingma, D. P. , & W. M. (2013a). *Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.*

Kingma, D. P. , & W. M. (2013b). *Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.*

Kingma, D. P. , & W. M. (2013c). *Kingma, D. P., & Welling, M. (2013, December). Auto-encoding variational bayes.*

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114. (2014).

Knoors, H. , & M. M. (2014). *Teaching deaf learners: Psychological and developmental foundations. OUP USA.*

Koch, G., Zemel, R., & Salakhutdinov, R. (2015, July). Siamese neural networks for one-shot image recognition. In ICML deep learning workshop (Vol. 2, No. 1, pp. 1-30).

Koller, O., Camgoz, N. C., Ney, H., & Bowden, R. (2019). Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos. IEEE transactions on pattern analysis and machine intelligence, 42(9), 2306-2320.

Koller, O., Camgoz, N. C., Ney, H., & Bowden, R. (2019). Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos. IEEE transactions on pattern analysis and machine intelligence, 42(9), 2306-2320.

Koller, O., Zargaran, S., Ney, H., & Bowden, R. (2018). Deep sign: Enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs. International Journal of Computer Vision, 126(12), 1311-1325.

Kopf, M., Schulder, M., & Hanke, T. (2022, June). The sign language dataset compendium: creating an overview of digital linguistic resources. In Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources (pp. 102-109).

Kostopoulou, K., Tholoniat, P., Cidon, A., Geambasu, R., & Lécuyer, M. (2023, October). Turbo: Effective caching in differentially-private databases. In Proceedings of the 29th Symposium on Operating Systems Principles (pp. 579-594).

Kotyan, A., Kirillov, V., & Lempitsky, V. (2024). On the Evaluation of Latent Spaces in Vision Models via k-Distributions. arXiv preprint arXiv:2408.09065. https://arxiv.org/abs/2408.09065*.

Kubicek, E., & Quandt, L. C. (2021). A positive relationship between sign language comprehension and mental rotation abilities. The Journal of Deaf Studies and Deaf Education, 26(1), 1-12. *The Journal of Deaf Studies and Deaf Education*.

Lahby, M., Maleh, Y., Bucchiarone, A., & Schaeffer, S. E. (2024). General Aspects of Applying Generative AI in Higher Education. Springer. https://doi. org/https://doi. org/10.1007/978-3-031-65691-0. (2024).

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474. (2020a).

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474. (2020b).

Li, P., Yan, H., & Lu, X. (2023). A Siamese neural network for learning the similarity metrics of linear features. International Journal of Geographical Information Science, 37(3), 684-711.

Liu, E., Lim, J. Y., MacDonald, B., & Ahn, H. S. (2024, August). Weighted Multi-modal Sign Language Recognition. In 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) (pp. 880-885). IEEE.

Liu, Y., Zhang, W., Ren, S., Huang, C., Yu, J., & Xu, L. (2025, April). SCOPE: Sign Language Contextual Processing with Embedding from LLMs. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 6, pp. 5739-5747).

Luckner, J., Bowen, S., & Carter, K. (2001). Visual teaching strategies for students who are deaf or hard of hearing. Teaching Exceptional Children, 33(3), 38-44.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.

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. (2023).

Lynn, P. (2004). The use of substitution in surveys. The Survey Statistician, 49(14-16), 211. (2004).

Madhiarasan, M., & Roy, P. P. (2022). A comprehensive review of sign language recognition: Different types, modalities, and datasets. arXiv preprint arXiv:2204.03328.

Marschark, M. , & H. P. C. (2012). H. deaf children learn: W. parents and teachers need to know. O. USA. (2012). *Marschark, M., & Hauser, P. C. (2012). How deaf children learn: What parents and teachers need to know. OUP USA.*

Mathieu, E., Rainforth, T., Siddharth, N., & Teh, Y. W. (2019, May). Disentangling disentanglement in variational autoencoders. In International conference on machine learning (pp. 4402-4412). PMLR.

Matsune, A., Hu, S., Li, G., Wen, S., Zhu, X., & Tan, Z. (2024). A geometry loss combination for 3d human pose estimation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3272-3281).

Meng, L., & Li, R. (2021). An attention-enhanced multi-scale and dual sign language recognition network based on a graph convolution network. Sensors, 21(4), 1120. (2021).

Miah, A. S. M., Hasan, M. A. M., Nishimura, S., & Shin, J. (2024). Sign language recognition using graph and general deep neural network based on large scale dataset. IEEE Access, 12, 34553-34569. (2024).

Mohandes, M., Deriche, M., & Liu, J. (2014). Image-based and sensor-based approaches to Arabic sign language recognition. IEEE transactions on human-machine systems, 44(4), 551-557.

Mokin, A., Sheshkus, A., & Arlazarov, V. L. (2025). Auto-Probabilistic Mining Method for Siamese Neural Network Training. Mathematics, 13(8), 1270. (2025).

Naresh, P. V., Visalakshi, R., & Satyanarayana, B. (2020). A Study on Sign Language Recognition-A Literature Survey. In ICDSMLA 2019: Proceedings of the 1st International Conference on Data Science, Machine Learning and Applications (pp. 745-752). Springer Singapore.

Nickolls, J., Buck, I., Garland, M., & Skadron, K. (2008). Scalable parallel programming with cuda: Is cuda the parallel programming model that application developers have been waiting for?. Queue, 6(2), 40-53. (2008).

Noor, T. H., Noor, A., Alharbi, A. F., Faisal, A., Alrashidi, R., Alsaedi, A. S., ... & Alsaeedi, A. (2024). Real-time arabic sign language recognition using a hybrid deep learning model. Sensors, 24(11), 3683.

Odaibo, S. (2019). *Odaibo, S. (2019). Tutorial: Deriving the standard variational autoencoder (vae) loss function. arXiv preprint arXiv:1907.08956.*

Oliphant, T. E. (2007). Python for scientific computing. Computing in science & engineering, 9(3), 10-20. (2007).

Othman, A. (2024). *Othman, A. (2024). Building Sign Language Datasets. In Sign Language Processing: From Gesture to Meaning (pp. 109-127). Cham: Springer Nature Switzerland.*

Owoc, M. L., Sawicka, A., & Weichbroth, P. (2019, August). Artificial intelligence technologies in education: benefits, challenges and strategies of implementation. In IFIP international workshop on artificial intelligence for knowledge management (pp. 37-58). Cham: Springer International Publishing. (2019).

Palanisamy, M., Mohanraj, R., Karthikeyan, A., & Mohanraj, E. (2024, December). SIGNEASE: AI-Driven American Sign Language Interpretation System. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 1670-1675). IEEE. (2024).

Palestinian Central Bureau of Statistics. (2022). *Palestinian Central Bureau of Statistics. (2022). Disability Survey in the State of Palestine. Ramallah, Palestine.*

Paszke, A. (2019). *Paszke, A. (2019). Pytorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703.*

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32. (2019).

Peralta, J. H. (2023). Microservice APIs: Using Python, Flask, FastAPI, OpenAPI and More. Simon and Schuster. (2023).

Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2015). Sign language recognition using convolutional neural networks. In Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13 (pp. 572-578). Springer International Publishing.

Pu, J., Zhou, W., Hu, H., & Li, H. (2020, October). Boosting continuous sign language recognition via cross modality augmentation. In Proceedings of the 28th ACM international conference on multimedia (pp. 1497-1505).

Rao, Y. S. N., Chong, Y. T., Khan, R. U., TEH, C. S., BARAWI, M. H., SUNAR, M. S., & SIM, J. J. J. (2024). Dynamic sign language recognition and translation through deep learning: A systematic literature review. Journal of Theoretical and Applied Information Technology, 102(21). (2024).

Raschka, S., Liu, Y. H., & Mirjalili, V. (2022). Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python. Packt Publishing Ltd.

Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. Expert Systems with Applications, 164, 113794.

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014, June). Stochastic backpropagation and approximate inference in deep generative models. In International conference on machine learning (pp. 1278-1286). PMLR.

Richards, M. (2015). Software Architecture Patterns. OReilly Media. Inc. (2015).

Rodriguez, M., Oubram, O., Bassam, A., Lakouari, N., & Tariq, R. (2025). Mexican Sign Language Recognition: Dataset Creation and Performance Evaluation Using MediaPipe and Machine Learning Techniques. Electronics, 14(7), 1423.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823). (2015).

Shaffer, D. W. (2005). Studio Mathematics: The Epistemology and Practice of Design Pedagogy as a Model for Mathematics Learning. WCER Working Paper No. 2005-3. Wisconsin Center for Education Research (NJ1). (2005).

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A re-view of the trends and challenges in adopting natural language processing methods for educa-tion feedback analysis. Ieee Access, 10, 56720-56739.). (2022).

Simonyan, K. , & Z. A. (2014). *Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.*

Singh, J. , & S. D. (2022). *Singh, J., & Singh, D. (2022, October). A comprehensive review on sign language recognition using machine learning. In 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-6). IEEE.*

Singha, J. , & D. K. (2013). *Singha, J., & Das, K. (2013). Recognition of Indian sign language in live video. arXiv preprint arXiv:1306.1301*.

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. Advances in neural information processing systems, 30. (2017).

(Sophokleous, A., Christodoulou, P., Doitsidis, L., & Chatzichristofis, S. A. (2021). Computer vision meets educational robotics. Electronics, 10(6), 730.). (2021).

Stinson, M., Gamta-Poddar, R., Meyer, L., Powers-Blom, C., & Singer, S. (2022). Effects of Messaging and Communication Strategy Training on Interaction in Teams

With Deaf and Hearing College Students. American Annals of the Deaf, 167(4), 431-456.

Sultana, R., Noreen, H., Khalid, H., Irshad, A., Sheikh, M., & ul Ain, Q. (2023). Common Communication Strategies Used by Teachers of Hearing-Impaired Children in Classroom Settings. Journal of Health and Rehabilitation Research, 3(1).

Tasyurek, S. M., Kiziltepe, T., & Keles, H. Y. (2025). Disentangle and regularize: Sign language production with articulator-based disentanglement and channel-aware regularization. arXiv preprint arXiv:2504.06610. (2025).

Trajkovski, G., & Hayes, H. (2025). The AI-Assisted Assessment Creation Framework. In AI-Assisted Assessment in Education: Transforming Assessment and Measuring Learning (pp. 59-114). Cham: Springer Nature Switzerland. (2025).

UNICEF. (2023). *UNICEF, Children with Disabilities in the State of Palestine, 2023.*

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Vilhelmsson, I. (2021). A performance comparison of an event-driven node. js web server and multi-threaded web servers. (2021).

Wang, F., Cheng, J., Liu, W., & Liu, H. (2018). Additive margin softmax for face verification. IEEE Signal Processing Letters, 25(7), 926-930. (2018).

Wang, J., Lin, Y., & Ma, A. J. (2020). Self-supervised learning using consistency regularization of spatio-temporal data augmentation for action recognition. arXiv preprint arXiv:2008.02086.

Wang, X., Chen, H., Tang, S. A., Wu, Z., & Zhu, W. (2024). Disentangled representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 9677-9696.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3-4), 129-164.

Xie, W., Chen, W., Shen, L., Duan, J., & Yang, M. (2021). Surrogate network-based sparseness hyper-parameter optimization for deep expression recognition. Pattern Recognition, 111, 107701.

Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. Annual review of sociology, 41(1), 331-357. (2015).

Xu, A., Hsieh, J. Y., Vundurthy, B., Cohen, E., Choset, H., & Li, L. (2022). Mathematical justification of hard negative mining via isometric approximation theorem. arXiv preprint arXiv:2210.11173. (2022).

Xu, M. , Y. S. , F. A. , & P. D. S. (2023). *Xu, M., Yoon, S., Fuentes, A., & Park, D. S. (2023). A comprehensive survey of image augmentation techniques for deep learning. Pattern Recognition, 137, 109347.*

Yang, C., Xu, Y., Dai, B., & Zhou, B. (2020). Video representation learning with visual tempo consistency. arXiv preprint arXiv:2006.15489. .

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610.

Zelinka, J. , & K. J. (2020). *Zelinka, J., & Kanis, J. (2020). Neural sign language synthesis: Words are our glosses. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 3395-3403).*

Zhang, B., Xu, H., Xiong, H., Sun, X., Shi, L., Fan, S., & Li, J. (2021). A spatiotemporal multi-feature extraction framework with space and channel based squeeze-and-excitation blocks for human activity recognition. Journal of Ambient Intelligence and Humanized Computing, 12(7), 7983-7995. (2021).

Zhang, D., Ke, S., Yang, J., & Anglin-Jaffe, H. (2024). Sign Language in d/Deaf Students' Spoken/Written Language Development: A Research Synthesis and Meta-

analysis of Cross-linguistic Correlation Coefficients. Review of Education, 12(3), E70016.

Zhang, H., Shen, C., Li, Y., Cao, Y., Liu, Y., & Yan, Y. (2019). Exploiting temporal consistency for real-time video depth estimation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1725-1734).

Zhao, R., Zhang, L., Fu, B., Hu, C., Su, J., & Chen, Y. (2024, March). Conditional variational autoencoder for sign language translation with cross-modal alignment. In Proceedings of the aaai conference on artificial intelligence (Vol. 38, No. 17, pp. 19643-19651).

Zhao, Z. (2023). Build a live news application with Next. js 13. (2023).

Zheng, J., Wang, Y., Tan, C., Li, S., Wang, G., Xia, J., ... & Li, S. Z. (2023). Cvt-slr: Contrastive visual-textual transformation for sign language recognition with variational alignment. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 23141-23150).

Zheng, Z. , & S. L. (2019). *Zheng, Z., & Sun, L. (2019). Disentangling latent space for vae by label relevant/irrelevant dimensions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12192-12201).*

Zuo, R., Potamias, R. A., Ververas, E., Deng, J., & Zafeiriou, S. (2024). Signs as Tokens: A retrieval-enhanced multilingual sign language generator. arXiv preprint arXiv:2411.17799.

هناء بنت عبدالله. (2021). مستوى التحديات التي تواجه ممارسات التنفيذ في البيئة التعليمية الرقمية لدى معلمي الصم وضعاف السمع & *,السالم, ماجد بن عبدالرحمن, الزهراني Level of the Challenges that Face Implementation Practices in the Digital Learning Environment for Teachers of Deaf and Hard of hearing Students. مجلة التربية الخاصة والتأهيل, 12(41.2), 197-254.*. (2021).

أسماء. (2022). ت. ا. ا. ع. ا. ف. ت. ا. ا. و. ا. ب. الثانوية. م. ك. ا. (أسيوط), 38(5. 2), & ,عبد العزيز الخضير 83-44. . (2022). *أسماء. (2022). تحديات التعلم القائم على المشاريع في تعليم* & *,عبد العزيز الخضير الطالبات الصم وضعاف السمع بالمرحلة الثانوية. مجلة كلية التربية (أسيوط), 38(5.2), 44-83.*.

Appendices

Appendix A. Additional Algorithms

Algorithm A.1: VAE Training Step

Inputs: batch x (B clips), VAE, weight β, λ_tc, λ_sep.

Outputs: losses L_rec, L_kl, L_tc, L_sep, L_total.

1. **Forward pass:$(\mu, \log\sigma^2, Z, \hat{X}) \leftarrow VAE(X)$.**
2. **Compute reconstruction loss L_rec.**
3. **Compute KL loss: $l\_kl=0.5\ \Sigma\ (\exp(\log\sigma^2\_i) + \mu\_i^2 - 1 - \log\sigma^2\_i)$.**
4. **Temporal separation loss L_sep.**
5. **Combine: $l\_total= l\_rec+\beta\ L\_kl + \lambda\_tc\ L\_tc + \lambda\_sep\ L\_sep$.**
6. **Backpropagate and update parameters.**
7. **Return all loss values.**

Algorithm A.2. Triplet Loss

Inputs: Embedding $E \in R^{\wedge}\ \{B \times d\}$, labels y, margin α

Outputs: Triplet loss L_trip

1. **Initialize $l\_trip=0$, $N\_trip=0$**
2. **For each anchor a:**
    a. **Identify positive p and negative N.**

b. **Choose the farthest positive p, compute d_ap.**

c. **Choose to closet negative n with d_an >d_ap.**

d. **If n exists: L_trip+= max (0, d_ap^2-d_an^2+α), increment N_trip**

3. **Return L_trip / max (1, N_trip).**

Algorithm A.3 1: Few-Shot Episodic Evaluation

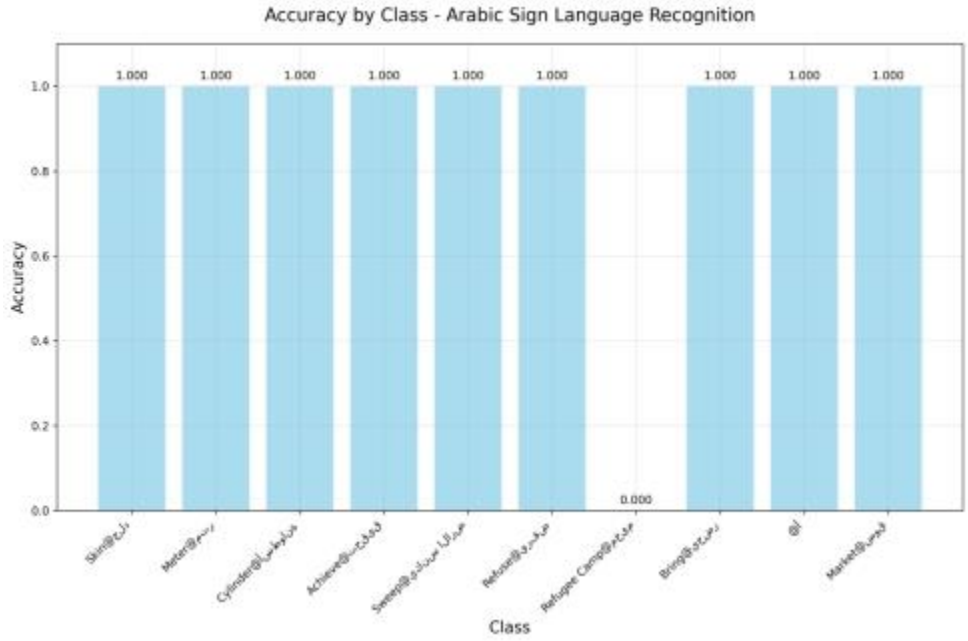
Inputs: Support set s, query set Q, encoder f

Outputs: Episode accuracy

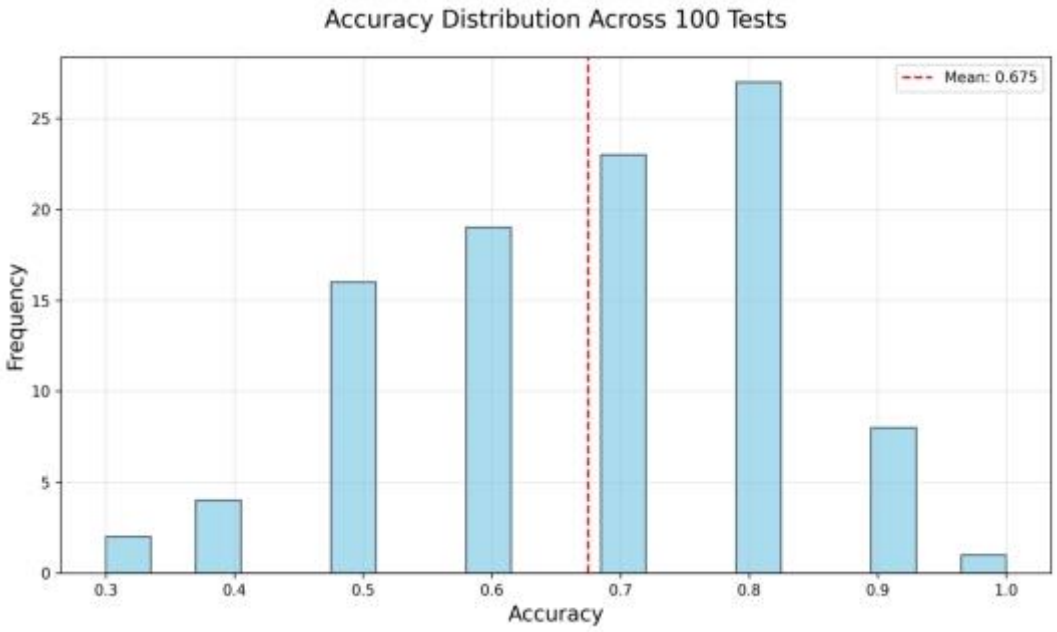
1. **For each class c: compute prototype = mean embeddings of support.**
2. **Initialize correct= 0, total = =|Q|.**
3. **For each query (x, y):**
   a. **Encode e= f(x).**
   b. **Predict class by nearest prototype.**
   c. **If predication y, increment correct.**
4. **Return correct/total.**

Appendix B. Additional Results

B.1: Accuracy by Class



B.2a. Accuracy Distribution Across 100 episodes with mean =0.675



B.2b. *Accuracy Distribution Across 100* episodes *with mean =0.641*

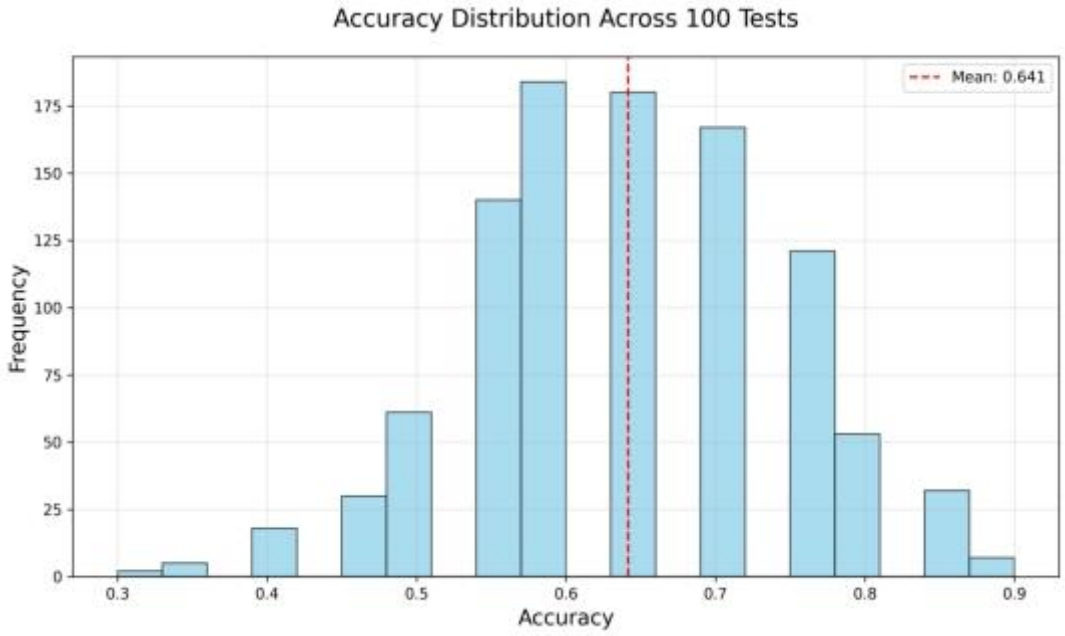
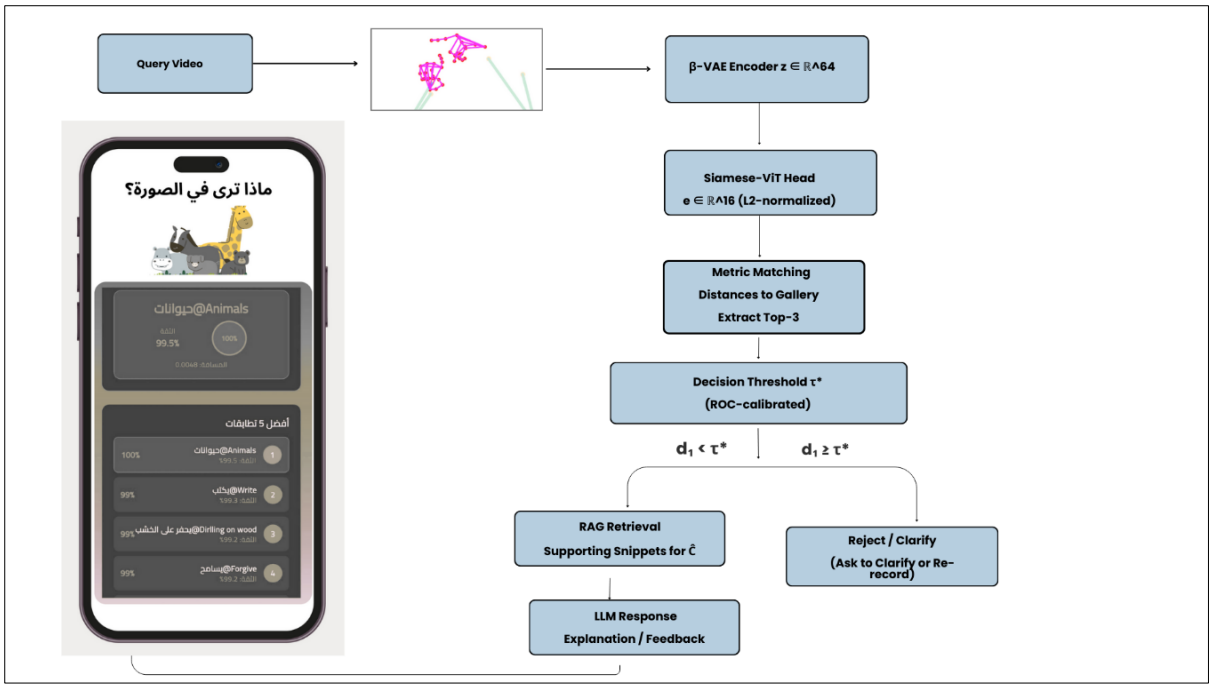


Table 4.1. Per Class Evaluation Results

Class (EN@AR)	(Precision %)	(Recall %)	F1 (%)	(Support)
Skin@جلد	100.0	100.0	100.0	1
Meter@متر	100.0	100.0	100.0	1
Cylinder@أسطوانة	100.0	100.0	100.0	1
Achieve@تحقيق	100.0	100.0	100.0	1
Sweep@ يكنس الارض	50.0	100.0	66.7	1
Refuse@يرفض	100.0	100.0	100.0	1
Refugee Camp@مخيم	0.0	0.0	0.0	1
Bring@يحضر	100.0	100.0	100.0	1

@أ	100.0	100.0	100.0	1
Market@سوق	100.0	100.0	100.0	1
Accuracy		90.0		10
Macro Avg	85.0	90.0	86.7	10
Weighted Avg	85.0	90.0	86.7	10

Figure 17.4. and Figure 18.4 Sign Pulse interface



Annex

List of experts for questioners

Name	Position	Institution
Dr. Amjad Shehadeh	Professor in educational technology	Birzeit university
Dr. Suher Alqasm	Director General of School Education	Ministry of Education
Mr. Majdi Moamar	Expert in STEM education	Ministry of Education
Eng. Laith Tura'ani	Expert in educational technology and STEM education	International private school
Mr. Khalil Alwneh	Expert in sign language	Ministry of Education
Ms. Nariman Sharwneh	Director General of Special Education	Ministry of Education
Ms. Nada Bazzar	Expert sign language	Palestine Red Crescent Society
Ms. Abeer	Expert sign language	Palestine Red Crescent Society
Dr. Ihab Shokri	Director General of Scientific Curricula	Ministry of Education
Dr. Naeem Koumi	Professor of mathematics	Arab American University
Mr. Ziad Sahloub	Expert in educational technology and STEM	Ministry of Education

Mr. Wailed Nazzal	Director General of Al-Amal Charitable Society for the Deaf	Al-Amal Charitable Society for the Deaf

Questionnaire review and judgment:

List of reviewers and judges:

Name	Title and position	Institution
Dr. Huthaifa Ashqar	Assistant Professor, lecturer	Arab American University
Dr. Abdelrahem Atawnih	Assistant Professor, Lecturer	Arab American University
Dr. Mohamed Khalil	Assistant Professor, Lecturer	Palestine Technical University – Kadoorie

Questionnaire

Expert questionnaires on the applicability and educational impact of the “Sign Pulse “ AI-based system for Palestinian sign language (PSL) recognition in STEM education.

I am Fidaa Khandaqji, a student currently pursuing my master’s degree in Artificial Intelligence at the Arab American University. I am preparing this questionnaire as part of the requirements for my master’s thesis under the supervision of Dr Huthaifa AL-Ashqar, titled: “Artificial intelligence system for Arabic sign language recognition to enhance education for deaf and hard-of-hearing students”. You are kindly requested to take about 15 minutes of your valuable time to complete this questionnaire. Please note that your responses will be used for research purposes only and will be treated with strict confidentiality.

Thank you very much for your kind cooperation and valuable time.

Fidaa Khandaqji

Title: Expert Questionnaire on the Applicability of an AI-based application for Palestinian Sign language in STEM Education.

Introduction to the application "SignPulse."

This master's thesis focuses on developing an artificial intelligence (AI) based application designed to recognize Palestinian sign language (PSL) and use it to support teaching STEM subjects for deaf and hard-of-hearing students in grades 1-4.

The application works by:

- Using computer vision to detect and recognize students' signs through a camera.
- Providing instant feedback and explanations related to the curriculum (math and science).
- Offering teachers and parents tools to monitor students' progress.

Purpose of the Questionnaire

The purpose of this questionnaire is to gather expert opinions on the applicability, benefits, challenges, and recommendations related to this proposed application. Your feedback will help evaluate its potential educational impact and guide future improvements. The questionnaire consists of four sections, as follows:

1. Section A: General information.
2. Section B: Evaluation of the applicability and usability of the system.
3. Section c: Assessment of the educational impact of the system
4. Section D: Experts' suggestions and recommendations for improvement.

Section A: General Information

1. Name (optional): _____

2. Specialization: ☐ special education
   ☐ Sign Language
   ☐ educational technology
   ☐ Others:_____

3. Years of experience: _____

Section B: Applicability

4. How important do you consider developing this application?

   ☐ Very important
   ☐ Important
   ☐ Moderately important
   ☐ Not important

5. What is the main benefit of such an application?

   ☐ Improving academic understanding
   ☐ Enhancing classroom interaction
   ☐ Reducing reliance on interpreters
   ☐ Other:_____

Section C: Challenges

6. What are the main challenges you expect?

   ☐ Lack of devices
   ☐ Need for teacher training
   ☐ Acceptance by students/ parents
   ☐ other:_____

Section D: Recommendation:

7. What suggestions would you provide to improve the Sign Pulse app?

   _____

8. Do you believe using AI in teaching DHH students is promising?
   ☐ Yes ☐ No
   If yes, please explain:_____

الملخص

تهدف هذه الدارسة الى تطوير وفاعيلة نظام sign pulse، وهو نظام تعليمي ذكي صمم لدعم الطلبة الصم وضعاف السمغ في تعلم مجالات العلوم والتكنولوجيا والهندسة والرياضيات (STEM) باستخدام لغة الاشارة الفلسطينية (PSL)، تعتمد الدارسة منهجية البحث المختلط ، حيث تجمع بين التحليلين النوعي والكمي لتقييم أثر النظام على تفاعل الطلبة ومستوى الفهم ، وجودة التفاعل.

تم جمع البيانات من خلال بيانات فيديو منظمة للغة الاشارة الفلسطينية، إلى جانب استبانية قابلة الاستخدام التي أكملها خبراء في لغة الإشارة ، وتكنولوجيا التعليم ، وتعليم STEM، وقد أتاحت هذه البيانات تدريب وتقييم نموذج هجين من نوع Siamese Vision Transformer (Siamese-ViT) لتحقيق تعلم دقيق للتعرف على الإشارات، في الوقت نفسه، وفر المكون التفاعلي للنظام، والمدعوم بتقنية RAG+GPT ، وشروحات مخصصة ، وأسئلة سياقية ، وتغذية راجعة فورية ، مما أسهم في تحسين تجربة التعلم.

تشير النتائج إلى دمج تقنيات الذكاء الاصطناعي يعزز بشكل ملحوظ إمكانية الوصول الى التعلم ، ومستوى التفاعل ، والفهم لدى الطلبة ذوي الإعاقات السمعية. كما تسلط الدارسة الضوء على الدور المحوري الإرشاد المتخصص والموارد الثقافية الملائمة للغة الإشارة في تعظيم فاعلية النظام، إضافة إلى ذلك، تناقش الدارسة التحديات المرتبطة بالقيود التقنية ، وتكيف المستخدمين ، ودمج النظام في الممارسات الصفية.

وخلصت الدارسة إلى أن أنظمة لغة الإشارة المعتمدة على الذكاء الاصطناعي ، مثل SignPulse، تمتلك إمكانيات كبيرة لتحسين التعليم الشامل لذوي الإعاقات السمعية، ودعم تطبيق استراتيجيات التعلم التكيفي، والمساهمة في توفير فرص تعليمية عادة ومنصفة في فلسطين.