Chapter 7
Evolving Approaches to Static American Sign Language Fingerspelling Recognition

Ajay Menon
https://orcid.org/0009-0004-7476-7842
School of Mathematical and Computer Sciences, Heriot Watt University, Dubai, UAE

Mahmoud A. A. Mousa
https://orcid.org/0009-0007-2415-5655
School of Mathematical and Computer Sciences, Heriot Watt University, Dubai, UAE

ABSTRACT

This review synthesizes research on static American Sign Language (ASL) alphabet recognition from images, comparing traditional machine learning pipelines, convolutional neural network (CNN) transfer learning, and hybrid or transformer-based models. The analysis spans 2016 to 2025 studies that detail preprocessing, model design, and quantitative results on datasets such as the ASL Alphabet and Sign Language MNIST. Classical approaches using engineered features with classifiers such as Support Vector Machines (SVMs) or Random Forests perform well in controlled settings but rely on robust segmentation and handcrafted descriptors. Transfer learning on CNN backbones, including MobileNetV2, ResNet, EfficientNet, DenseNet, and the Visual Geometry Group (VGG) models, achieves near-perfect within-dataset accuracy; pure and modified Vision Transformers (ViTs)

DOI: 10.4018/979-8-3373-2033-5.ch007

and CNN–transformer hybrids also reach ceiling-level performance with favorable speed-to-accuracy tradeoffs. Most evaluations remain closed set and seldom report signer-independent splits, cross-dataset transfer, or deployment metrics.

INTRODUCTION

Hearing loss is a large and growing public health challenge with direct implications for communication access and equity. The World Health Organization (WHO) reports that more than 5% of the world's population, or about 430 million people, require rehabilitation for disabling hearing loss. By 2050, nearly 2.5 billion individuals are projected to be affected, including more than 700 million who will need rehabilitation. The annual global cost of unaddressed hearing loss approaches one trillion US dollars. In addition, over one billion young adults remain at risk due to unsafe listening practices (WHO, 2025). In parallel, disability affects an estimated 1.3 billion people worldwide, or roughly one in six, underscoring the systemic nature of access barriers in health, education, and employment (WHO, 2023b).

For deaf communities, timely access to information and services in sign language remains uneven despite international commitments to inclusion. The World Federation of the Deaf (WFD) reports that only about 41.5 percent of countries legally recognize their national sign languages, even as the global deaf population is often cited at roughly 70 million across more than 200 distinct sign languages (Dsane, 2024; WFD, 2025). Accessibility is integral to health information ecosystems and public communication, not peripheral accommodation, which places a premium on technologies that can bridge language and modality gaps at scale (WHO, 2023a). Within this context, automated recognition of signed alphabets offers a pragmatic entry point for communication aids, educational tools, and user interfaces, if models are accurate, efficient, and robust beyond laboratory conditions.

This review focuses on a small but important part of the landscape: static American Sign Language (ASL) alphabet recognition from images. The scope excludes temporal modeling of dynamic letters and continuous signing to examine, with greater depth, how current methods handle data curation and preprocessing, feature extraction, architecture choice, training protocols, and evaluation on static benchmarks. The aim is to synthesize what works, where, and at what computational cost, while making explicit the assumptions that underline near-ceiling results on curated datasets. The specific objectives are to compare traditional machine learning pipelines against transfer-learned convolutional neural network (CNN) models, hybrid CNN–transformer designs, and pure and modified transformer approaches; to align reported accuracies with dataset characteristics and splits; to identify gaps in signer diversity, cross-dataset validation, fairness, and deployment metrics; and