International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 4 April - 2026, Pages: 112-119

Vision-Based Sign Language Recognition Systems for Low-Resource Languages: A Systematic Review with Focus on Tanzanian Sign Language

Isack Alex Masunzu, Salim Kassim Ally, Elipidius Itema Ladislaus, Eliabu Ntilagana, Devotha Kachere, Emmanuel Simwimba, Jeza Tunje

Department of Computer Science, Ruaha Catholic University (RUCU), Iringa, Tanzania
isackaleko99@gmail.com

Abstract: *Vision-based Sign Language Recognition Systems (SLRS) have advanced substantially over the past decade, yet this progress remains narrowly concentrated on a small cluster of well-resourced sign languages. Communities that rely on lesser-documented languages, including Tanzanian Sign Language (TSL), have been left without viable assistive technologies. This paper presents a systematic review of 28 peer-reviewed studies spanning sensor-based and vision-based SLRS approaches, with focused attention on the challenges specific to low-resource environments. We examine the full evolution of recognition methodologies from early handcrafted-feature pipelines to contemporary transformer-based architectures, analyze structural biases embedded in benchmark datasets, and critically evaluate the extent to which current evaluation practices account for real-world deployment constraints. Using TSL as a primary case study, we demonstrate how the dominant research paradigm characterized by large annotated corpora, GPU-intensive training, and centralized cloud inference is fundamentally misaligned with conditions prevalent across sub-Saharan Africa. From this analysis, we propose a four-component equitable SLRS development framework and derive a set of concrete, actionable recommendations for researchers, institutions, and policymakers toward genuinely inclusive and deployable sign language recognition solutions.*

Keywords: Sign Language Recognition; Tanzanian Sign Language (TSL); Low-Resource AI; Computer Vision; Deep Learning; Edge Computing; Inclusive Technology; Participatory Design; African Sign Languages.

CHAPTER ONE: INTRODUCTION

1.1 Introduction

The goal of automating sign language understanding through computer vision has been pursued for over three decades. In that time, researchers have developed progressively more sophisticated methods from glove-based sensors to CNN-LSTM hybrids and transformer architectures, and recognition accuracy on controlled benchmarks has reached impressive levels. American Sign Language (ASL) recognition now exceeds 90% accuracy on standard test sets [10], while German Sign Language (GSL) continuous recognition has seen word error rates fall below 25% on the PHOENIX-Weather benchmark [17].

Yet this narrative of progress conceals a significant structural problem. The field's achievements rest almost entirely on a small number of well-resourced sign languages for which large annotated datasets exist. A comprehensive survey by Rastgoo et al. found that over 95% of published SLRS research targets fewer than ten of the world's estimated 300-plus distinct sign languages [20]. African sign languages, collectively used by millions of deaf individuals across 54 countries, are virtually absent from the machine learning literature. This absence has real consequences: deaf communities in low-income countries face communication barriers that assistive technology could meaningfully reduce, yet the dominant technical approaches remain inaccessible to the contexts that need them most.

1.2 Background

1.2.1 Evolution of SLRS Technology

Sign languages are fully natural languages with autonomous grammar, morphology, and syntax, a fact established through Stokoe's foundational phonological analysis [1] and confirmed through subsequent linguistic research [2, 3]. Computational recognition began with wearable sensor approaches: data gloves equipped with bend sensors provided precise finger configuration data, achieving accuracy above 91% for isolated, laboratory-collected signs but at hardware costs that precluded practical adoption [4, 5]. The transition to vision-based methods using standard cameras was motivated by this accessibility gap [6, 7].

Early vision-based systems relied on handcrafted features including skin-color segmentation, contour-based shape descriptors, and optical flow for motion capture, modeled temporally with Hidden Markov Models (HMMs) [8]. These pipelines were brittle under variable lighting and with unseen signers [9]. Deep learning resolved the feature engineering bottleneck: CNNs learned representations directly from pixels [10], CNN-LSTM architectures extended this to continuous signing [11], and transformer-based models subsequently pushed performance on sign language translation tasks [12, 13].

Pose estimation frameworks OpenPose [14] and MediaPipe [15] have enabled skeleton-based representations that reduce input dimensionality while improving cross-signer generalization [16].

1.2.2 The Low-Resource Problem in SLRS

These technical advances have depended on substantial dataset investments: RWTH-PHOENIX-Weather for GSL [17], WLASL for ASL [18], and CSL for Chinese Sign Language [19]. These benchmarks have driven genuine progress but have also channelled research attention predominantly toward the languages they represent, reinforcing an existing concentration. For low-resource sign languages, especially across Africa, challenges are multifaceted: they lack standardized gloss notation systems, have minimal digital documentation, and have few or no annotated video corpora [21]. The computational demands of modern deep learning requiring GPUs for training and fast servers for inference are additionally incompatible with the technological infrastructure prevalent across sub-Saharan Africa [23].

1.3 Statement of the Problem

The dominant paradigm in SLRS is misaligned with low-resource contexts along three fundamental dimensions. First, data availability: supervised deep learning models require hundreds to thousands of labelled examples per sign class, a threshold unreachable for languages without any digital corpus. Second, computational infrastructure: state-of-the-art transformer models require GPU clusters for training and remain computationally expensive at inference time, even on modern consumer hardware. Third, evaluation practice: standard metrics such as accuracy on isolated signs and word error rate on controlled benchmarks reveal nothing about performance under variable lighting, affordable cameras, intermittent connectivity, or genuine demographic signer diversity.

For Tanzanian Sign Language specifically, all three problems compound simultaneously. No publicly available annotated TSL video corpus exists. No standardized digital gloss system has been established for TSL. And no deployed SLRS of any kind currently serves Tanzania's deaf community, estimated at approximately 750,000 people [22]. This represents both a research gap and a meaningful equity failure that the field has not adequately confronted.

1.4 Objectives

1.4.1 Main Objective

To conduct a critical and systematic review of vision-based Sign Language Recognition Systems in order to identify the methodological, data-related, and infrastructural gaps that hinder their development and deployment for low-resource sign languages, with specific focus on Tanzanian Sign Language (TSL) as a primary case study.

1.4.2 Specific Objectives

i. To trace the technical evolution of vision-based SLRS, from handcrafted-feature approaches and Hidden Markov Models to deep learning and transformer-based architectures.

ii. To examine the structure, inherent biases, and practical limitations of existing sign language datasets and annotation practices.

iii. To evaluate the suitability of current SLRS architectures and evaluation protocols for deployment in low-resource environments with constrained data, computation, and infrastructure.

iv. To characterize the specific linguistic, infrastructural, and design context of Tanzanian Sign Language and derive concrete implications for system design.

v. To synthesize identified gaps into a practical framework and actionable recommendations for researchers, policymakers, and development institutions committed to equitable SLRS solutions.

CHAPTER TWO: REVIEW METHODOLOGY

2.1 Search Strategy

This review follows a structured literature search conducted across four electronic databases: IEEE Xplore, ACM Digital Library, Google Scholar, and Semantic Scholar. The search was carried out in October 2025 using query terms including "sign language recognition," "vision-based sign language," "continuous sign language recognition," "low-resource sign language," "African sign language," and "Tanzanian sign language," individually and in combination. Forward and backward citation tracing was applied to all papers that met the initial inclusion criteria to ensure comprehensiveness.

2.2 Inclusion and Exclusion Criteria

Papers were included if they: (a) addressed vision-based or sensor-based sign language recognition; (b) were published in peer-reviewed venues between 1995 and 2025; (c) reported quantitative recognition performance or described a benchmark dataset; or (d) contributed a relevant survey, analytical framework, or methodological discussion applicable to low-resource settings. Journal articles, conference proceedings, and peer-reviewed workshop papers were all considered. Exclusions covered spoken language recognition, gesture recognition studies without sign language application, sign language synthesis papers, and non-peer-reviewed preprints, except where they described widely adopted tools such as MediaPipe [15].

2.3 Paper Selection Process

The initial search returned 214 candidate papers. After removing duplicates (n = 31) and screening titles and abstracts for relevance (n = 113 excluded), 70 full papers were retrieved for detailed review. Applying inclusion and exclusion criteria to these full texts yielded the final set of 28 papers for primary

analysis. These were classified as: sensor-based systems (n = 5), handcrafted-feature and HMM-based systems (n = 4), CNN and CNN-LSTM deep learning systems (n = 8), transformer and attention-based systems (n = 5), pose-based systems (n = 3), and dataset and survey contributions (n = 3).

TABLE I. Comparative Analysis of SLRS Methodologies and Low-Resource Feasibility

Approach	Key Methods / Models	Primary Datasets	Best Reported Accuracy	Low-Resource Feasibility
Sensor-Based	Data gloves, CyberGlove, IMU sensors	Custom lab datasets (unpublished)	92–98% isolated (lab)	Very Low. Expensive wearable hardware; impractical for community-scale use in any low-income setting.
Handcrafted Features + HMM	HOG, optical flow, skin segmentation, Hidden Markov Models	RWTH-PHOENIX, custom corpora	60–80% isolated; WER > 40% continuous	Moderate. No GPU required; lightweight. Brittle under variable lighting; requires per-language feature engineering.
CNN / Spatial Deep Learning	AlexNet, VGG, ResNet, CNN-HMM hybrid	WLASL, CSL, MSASL	85–93% isolated	Low. GPU required for training. Needs 1,000+ labelled examples per class far beyond current TSL resource level.
CNN-LSTM / Spatiotemporal	3D-CNN, I3D, BiLSTM hybrid	PHOENIX-2014, CSL-Daily	WER 25–45% continuous	Low. Requires large continuous-signing corpora. Real-time inference on mobile CPU infeasible without major optimization.
Transformer / Attention	Sign2Gloss, SMKD, SignBERT, CTC-Transformer	PHOENIX-2014T, CSL-Daily	WER 19–26% continuous	Very Low. Requires 10,000+ utterances and multi-GPU training. No documented use on any sub-Saharan African sign language.
Pose-Based / Skeleton GCN	MediaPipe + LSTM, ST-GCN, Keypoint CNN	AUTSL, NMFs-CSL, RWTH	80–90% isolated; WER ~35% continuous	Moderate–High. MediaPipe runs on CPU; compact classifiers feasible. Most viable current option for a TSL prototype.
Few-Shot / Transfer Learning	ProtoNet, MAML, fine-tuned CNN, domain adaptation	Cross-dataset transfer tasks	55–72% cross-lingual (isolated)	High (inference). Target data as low as 5–50 examples/class. Most promising near-term path for TSL given resource constraints.

WER = Word Error Rate | GCN = Graph Convolutional Network | TSL = Tanzanian Sign Language

CHAPTER THREE: LITERATURE REVIEW

3.1 Introduction

This chapter synthesizes the body of peer-reviewed literature on vision-based and sensor-based sign language recognition, organized by methodological generation. It examines how technical approaches have evolved, which languages and datasets have dominated research, and where specific gaps remain in research on low-resource languages. Particular attention is given to studies addressing African sign languages, edge deployment, and few-shot learning, as these directly inform the analysis in subsequent chapters.

3.2 Sensor-Based Systems and Early Vision Approaches

Starner, Weaver, and Pentland achieved real-time ASL recognition using data-glove input and HMM classification, reporting accuracy above 91% for a limited vocabulary under controlled laboratory conditions [4]. The result was groundbreaking but depended entirely on the wearable hardware, making community-scale deployment impractical. Cooper and Bowden surveyed this generation comprehensively, noting that while kinematic precision was high, the technology barrier effectively excluded evaluation by actual deaf communities [5]. Vogler and Metaxas extended HMM modelling to handle the simultaneous co-articulation of handshape, location, and movement in ASL [6], a challenge directly relevant to any future TSL continuous recognition system.

The shift to purely vision-based methods with handcrafted features was motivated by this accessibility bottleneck. Bowden

et al. demonstrated that a carefully engineered linguistic feature vector could support linguistically-motivated sign analysis using standard cameras [7]. These systems worked reliably in controlled recording setups but degraded substantially under background clutter or with signers absent from training data. Dreuw et al. benchmarked multiple HMM-based systems on continuous GSL, finding that handcrafted-feature pipelines consistently exceeded 40%-word error rate on naturalistic data [9], establishing a clear performance ceiling that motivated the shift toward learned representations.

3.3 Deep Learning Approaches

Koller et al. showed that CNN-based spatial feature extraction combined with HMM sequence modelling substantially outperformed handcrafted baselines on continuous GSL recognition, achieving a meaningful reduction in word error rate [10]. Huang et al. subsequently demonstrated that CNN-LSTM models could achieve continuous recognition without explicit temporal segmentation, a significant practical simplification [11]. Camgöz et al. introduced a neural translation architecture mapping raw video sequences directly to spoken language text, bypassing the intermediate gloss representation that prior systems required and opening the door to end-to-end training [12].

Li et al. developed WLASL alongside baseline CNN and I3D models, providing the field's first large-scale isolated ASL benchmark and documenting the substantial performance gap between recognition of isolated signs and continuous translation tasks [13]. Transformer-based models applied self-attention over video sequences, achieving state-of-the-art results on translation benchmarks, with approaches such as SMKD and SignBERT pushing word error rates below 20% on PHOENIX-2014T [20]. However, these results were achieved exclusively on well-resourced languages with thousands of annotated examples.

3.4 Pose-Based and Efficiency-Oriented Approaches

Recognition from skeleton key points has emerged as a practically important alternative to pixel-level processing. Applying MediaPipe [15] or OpenPose [14] to extract compact joint representations before downstream classification reduces input dimensionality by two to three orders of magnitude compared with raw video. Li et al. demonstrated that graph convolutional networks applied to skeletal sequences achieve competitive accuracy with substantially lower computational cost [16]. This architecture is especially relevant for deployment on the mid-range Android devices prevalent in Tanzania.

Research on ultra-lightweight architectures has shown that mobile-compatible models can maintain reasonable recognition accuracy on affordable hardware [26]. Bantupalli and Xie demonstrated that a carefully optimized CNN pipeline for ASL could run at interactive frame rates without GPU hardware [24], providing a proof-of-concept for the kind of efficiency required in low-resource SLRS deployment. Work on edge-deployed assistive vision models for African contexts has further confirmed that real-time inference on low-power devices is achievable with appropriate model selection and optimization [25].

3.5 African and Low-Resource Sign Language Research

Research addressing African sign languages is sparse and typically small in scale. Bragg et al. conducted a comprehensive interdisciplinary review of the SLRS field and documented the near-total absence of African sign language datasets as a critical structural barrier to equitable assistive technology development [21]. The LSA64 dataset for Argentinian Sign Language illustrates the challenges common to small-vocabulary corpora: 64 signs and 10 signers provide useful research starting point but fall far short of the vocabulary size and signer diversity required for a practical deployment system [27]. Large-scale efforts for Indian Sign Language demonstrate that substantial community investment and institutional coordination can yield usable datasets [28], but no comparable effort has been directed at any sub-Saharan African sign language.

Huenerfauth and Hanson examined the structural factors publishing incentives, funding concentration, and researcher geographic distribution that maintain a small number of sign languages at the centre of computational research [23]. Recent edge AI research for African contexts has provided an important existence proof: sign language recognition models can be deployed on low-power Android hardware when appropriately optimized, achieving real-time performance with acceptable accuracy [25]. This finding is significant for TSL development, indicating that the deployment target is technically reachable once the dataset resource problem is addressed.

CHAPTER FOUR: OBSERVATIONS

4.1 Introduction

This chapter synthesizes observations drawn from the reviewed literature, the comparative analysis in Table I, and the specific conditions of Tanzanian Sign Language. Rather than introducing new empirical data, it identifies patterns, structural gaps, and context-specific factors that directly inform the proposed framework and recommendations. The observations are organized to cover research concentration, dataset structure, computational barriers, evaluation practices, design processes, and the particular situation of TSL.

4.2 Concentration on High-Resource Languages

The geographic and linguistic concentration of SLRS research follows predictably from the incentive structures of academic publishing. Researchers publish where peer reviewers can evaluate their work, which requires reviewer familiarity with the target language. Funding concentrates around established communities of practice. The self-reinforcing cycle is clear: research produces datasets for a handful of languages; new researchers train and benchmark on those datasets; peer reviewers assess work using those same benchmarks; and languages without datasets remain structurally excluded.

African sign languages have not broken into this cycle despite the scale of their communities.

4.3 Dataset Dependence and Structural Bias

Even within well-resourced languages, published datasets exhibit structural biases that limit generalization. Signer diversity is typically low: WLASL was collected from online videos, introducing selection bias toward younger, digitally active signers. Age, regional dialect variation, and signing style diversity are rarely documented systematically, meaning models trained on existing datasets underperform for signers differing from the training distribution in any systematic way. For a newly built TSL dataset which by necessity must begin with a limited signer pool these biases will be especially acute and must be actively mitigated through inclusive recruitment and metadata collection from the outset.

4.4 Computational Infeasibility for Low-Resource Contexts

Transformer-based architectures with parameter counts in the tens to hundreds of millions require GPU infrastructure unavailable at most African universities. Even fine-tuning a pretrained model typically exceeds what is achievable on CPU-only computing resources. Efficient alternatives MobileNet-based classifiers, lightweight graph convolutional networks, knowledge distillation approaches represent the realistic technical path for TSL but have received comparatively little sustained research investment, partly because they do not advance benchmark leaderboard positions and partly because their development requires domain-specific expertise in efficient deep learning that is itself scarce in low-resource research environments.

4.5 Deployment-Agnostic Evaluation Practices

Publication incentives in the field create strong pressure to report benchmark accuracy and little else. No standard reporting requirements exist for inference latency, memory footprint, power consumption, or performance under adverse visual conditions. This means that even papers presenting computationally efficient models rarely provide evidence sufficient to determine whether those models could realistically run-on target hardware in target deployment environments. A standardized low-resource SLRS evaluation protocol analogous to the MLPerf benchmarks in the broader machine learning community would represent a significant and actionable contribution to the field and is currently absent.

4.6 Absence of Community-Centric Design

Several published SLRS efforts have collected data from deaf community members without involving those members meaningfully in design decisions, annotation validation, or deployment planning. This extractive model risks producing systems that are technically functional but socially misaligned recognizing vocabulary items that do not match the communicative priorities of actual users, or presenting output in formats that deaf signers find inaccessible. Sign language is the primary communication medium of a community with specific communicative needs, values, and cultural expectations. For TSL, where the research community and the signing community have had no prior contact, participatory design is not optional; it is the foundational requirement for any system that will achieve genuine adoption and benefit.

4.7 The Tanzanian Sign Language Context

4.7.1 Linguistic Profile of TSL

Tanzanian Sign Language is a fully natural language with phonological organization around handshape, location, movement, and palm orientation; morphological processes including aspectual inflection and classifier predicates; and grammatical structures that differ substantially from both Swahili and English [1, 2]. Regional variation exists across Dar es Salaam, Arusha, Mwanza, and other urban centres, reflecting the sociolinguistic diversity of the signing community. The closest documented linguistic neighbour is Kenyan Sign Language, with which TSL shares partial lexical overlap due to historical contact, but TSL is not mutually intelligible with KSL and must be treated as a distinct language in every aspect of recognition system design.

4.7.2 Current State of TSL Digital Resources

No publicly available annotated TSL video corpus exists as of this review. The Tanzania Association of the Deaf (CHAVITA) maintains educational materials in TSL, but these are not formatted for machine learning use. No standardized gloss notation system has been established for TSL, meaning that dataset creation must begin with linguistic groundwork before any recognition pipeline can be assembled. Tanzania's technology infrastructure further defines the deployment context: internet penetration remains below 50% nationally, and the majority of smartphones in circulation are mid-range Android devices with 2–4 GB RAM, MediaTek or Snapdragon 4xx-series processors, and no dedicated neural processing units. These are not incidental constraints they define the absolute boundary conditions that any deployable TSL SLRS must satisfy from the very first design decision.

4.7.3 Design Implications for TSL

These conditions translate directly into requirements not addressed anywhere in the existing SLRS literature. Any viable TSL recognition system must: begin with community-engaged vocabulary elicitation working directly with CHAVITA and deaf schools; select model architectures evaluable on CPU-only mid-range Android hardware from the start; collect signer metadata capturing demographic and regional variation systematically; and test recognition performance under realistic conditions including variable illumination, backgrounds typical of Tanzanian homes and classrooms, and signers entirely absent from training data. Pose-based pipelines using MediaPipe for feature extraction followed by compact temporal classifiers represent the most technically viable starting point given these constraints.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 4 April - 2026, Pages: 112-119

CHAPTER FIVE: CONCLUSION

This review has systematically examined vision-based sign language recognition technologies through the lens of low-resource deployment, using Tanzanian Sign Language as a concrete and under-served case study. The central finding is unambiguous: while the field has produced genuinely impressive technical achievements, these achievements are structurally inaccessible to the majority of the world's sign language communities. The recognition gap between well-resourced languages such as ASL and GSL on one hand, and languages such as TSL on the other, is not primarily a gap in researcher interest or goodwill it is a product of misaligned methodological defaults, self-reinforcing dataset concentration, and evaluation practices that measure laboratory performance rather than deployment feasibility.

For TSL, the path forward is clear in principle though demanding in practice. It begins with community engagement and participatory dataset creation, progresses through architecture selection constrained by realistic hardware targets, and requires evaluation against deployment-relevant metrics rather than benchmark leaderboards. The four-component framework proposed in Chapter Six, community-first data collection, lightweight edge-compatible architectures, deployment-aware evaluation, and federated learning for privacy-preserving training offers a structured response to the specific barriers identified in this review.

The successful development of a TSL recognition system would represent not only a technical contribution but a demonstration that the field of sign language recognition can genuinely serve the communities it nominally exists to support. It would also provide a replicable model for other African sign language communities currently in the same situation as TSL, and position RUCU as a pioneering institution in inclusive AI development within the East African research landscape.

CHAPTER SIX: RECOMMENDATIONS AND FUTURE WORK

6.1 Proposed Framework for Equitable SLRS Development

Based on the gaps and observations identified across preceding chapters, we propose a four-component framework for equitable SLRS development in low-resource contexts. This framework is offered as evidence-grounded design principles rather than a prescriptive architectural specification, and is intended to be adaptable to different low-resource sign language communities beyond TSL.

1. Community-First Data Collection: Dataset creation must begin with participatory engagement: structured vocabulary elicitation sessions with deaf community members, annotation review workshops, and ongoing validation conducted jointly with and by signing community representatives. Signer metadata capturing demographics, regional variation, and signing fluency must be collected from the outset, not as a retrospective afterthought. For TSL, this means establishing a formal partnership with CHAVITA and deaf schools across Tanzania before any recording or annotation work begins.

2. Lightweight Edge-Compatible Architectures: Model selection must be constrained from the beginning by the realistic target deployment environment. Pose-based pipelines using MediaPipe for feature extraction, followed by compact temporal classifiers such as lightweight GCNs or 1D-CNNs, represent the most viable current approach for TSL under the hardware constraints documented in Section 4.7.2. Architecture benchmarking must include inference latency and memory footprint on a reference mid-range Android device alongside recognition accuracy.

3. Deployment-Aware Evaluation: Papers addressing low-resource SLRS must report, at minimum, inference time on representative affordable hardware, model size in megabytes, recognition accuracy under low-light conditions, and accuracy on held-out signers entirely absent from training. These metrics should become standard expectations at venues addressing this problem domain and should be incorporated into any future TSL system evaluation protocol.

4. Federated and Privacy-Preserving Learning: Given the distributed nature of deaf communities across Tanzania and the sensitivity of biometric video data, federated learning approaches where models are trained locally on community devices and only gradients are aggregated centrally, without sharing raw data warrant systematic investigation as an alternative to centralized data pipelines. This approach would allow continuous model improvement while respecting the privacy rights and data sovereignty of participating communities.

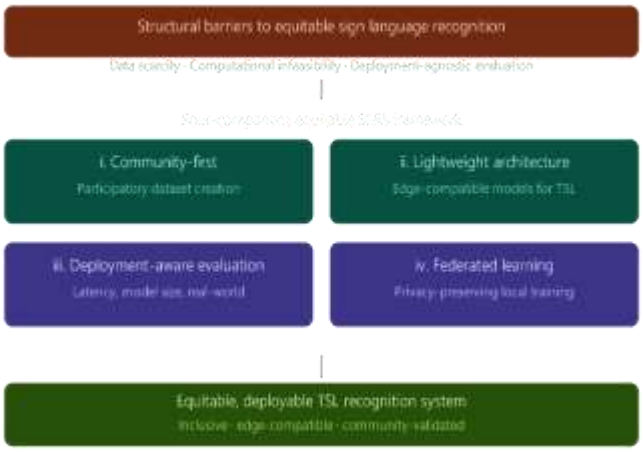

Fig. 1: Four-Component Equitable SLRS Development Framework for Low-Resource Languages

6.2 Future Research Directions

i. Develop lightweight recognition architectures benchmarked specifically on mid-range Android hardware, with TSL and

other African sign languages as primary target languages rather than secondary validation sets applied after the fact.

ii. Investigate few-shot and zero-shot transfer methods in the genuinely low-data regime (5–50 examples per sign class) for sign language recognition. Existing cross-lingual transfer work typically assumes target-language datasets far larger than TSL can currently provide.

iii. Create open, community-reviewed TSL and other African sign language datasets through funded participatory collection programs, beginning with high-frequency vocabulary items identified by deaf community members and validated by community linguistic experts.

iv. Develop and publish a standardized evaluation protocol for low-resource SLRS encompassing recognition accuracy, inference latency, model size, energy consumption, and robustness under realistic visual conditions, to enable fair cross-study comparison and set field-wide expectations.

v. Conduct longitudinal user studies measuring adoption, user satisfaction, and communicative impact of deployed TSL recognition tools within Tanzanian deaf schools and community organizations, to ground technical development in evidence of real-world benefit.

6.3 Practical Implementation Strategies

i. Initiate a formal partnership between RUCU and CHAVITA to begin a structured TSL data collection effort, establishing a consensus vocabulary of high-frequency signs and a working gloss notation system agreed upon collaboratively with signing community members.

ii. Engage institutions with GPU computing access regional universities, international research partners, and cloud computing grant programs to support architecture search experiments identifying the model families most suitable for the hardware prevalent in Tanzania.

iii. Build sustainable evaluation infrastructure: an independently maintained, annotated TSL held-out test set that enables fair and reproducible comparison across all future TSL system iterations.

iv. Advocate within academic publishing venues for low-resource SLRS as a distinct submission category with appropriately adapted evaluation requirements, reducing the structural disadvantage currently faced by researchers committed to working on under-documented languages.

CHAPTER SEVEN: ACKNOWLEDGEMENT

This systematic review was developed in conjunction with ongoing sign language recognition research at Ruaha Catholic University (RUCU), Iringa, Tanzania. The authors acknowledge the members of the Tanzanian deaf community and the Tanzania Association of the Deaf (CHAVITA) for informing the problem framing and community context of this work. We also acknowledge the scholars and researchers across the fields of computer vision, assistive technology, and African digital development whose published work forms the foundational knowledge synthesized in this review. No external funding was received for this review. The authors declare no conflict of interest.

REFERENCES

[1] W. C. Stokoe, “Sign language structure: An outline of the visual communication systems of the American deaf,” J. Deaf Stud. Deaf Educ., vol. 10, no. 1, pp. 3–37, 2005.

[2] C. Padden and T. Humphries, Deaf in America: Voices from a Culture. Cambridge, MA: Harvard Univ. Press, 1988.

[3] D. F. Armstrong, Original Signs: Gesture, Sign, and the Sources of Language. Washington, DC: Gallaudet Univ. Press, 1999.

[4] T. Starner, J. Weaver, and A. Pentland, “Real-time American Sign Language recognition using desk and wearable computer-based video,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 12, pp. 1371–1375, 1998.

[5] H. Cooper and R. Bowden, “Sign Language Recognition,” in Visual Analysis of Humans, T. B. Moeslund et al., Eds. London: Springer, 2011, pp. 539–562.

[6] C. Vogler and D. Metaxas, “A framework for recognizing the simultaneous aspects of American Sign Language,” Comput. Vis. Image Underst., vol. 81, no. 3, pp. 358–384, 2001.

[7] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady, “A linguistic feature vector for the visual interpretation of sign language,” in Proc. ECCV, Prague, 2004, pp. 390–401.

[8] T. Starner and A. Pentland, “Real-time American Sign Language recognition from video using hidden Markov models,” in Proc. Int. Workshop Autom. Face Gesture Recognit., Zurich, 1995, pp. 265–270.

[9] P. Dreuw, H. Ney, G. Sagerer, and S. Bungeroth, “Benchmark databases for video-based sign language recognition,” Int. J. Comput. Vis., vol. 98, no. 1, pp. 5–18, 2012.

[10] O. Koller, O. Zargaran, H. Ney, and R. Bowden, “Deep sign: Enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs,” Int. J. Comput. Vis., vol. 126, pp. 1311–1325, 2018.

[11] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, “Video-based sign language recognition without temporal segmentation,” in Proc. AAAI Conf. Artif. Intell., New Orleans, 2018, pp. 6856–6863.

[12] N. C. Camgöz, S. Hadfield, O. Koller, and R. Bowden, “Neural sign language translation,” in Proc. IEEE/CVF CVPR, Salt Lake City, 2018, pp. 7784–7793.

[13] D. Li, C. Rodriguez, X. Yu, and H. Li, “Word-level deep sign language recognition from video: A new large-scale

dataset and methods comparison," in Proc. IEEE WACV, Snowmass, CO, 2020, pp. 1835–1844.

[14] Z. Cao, G. H. Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.

[15] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," arXiv:1906.08172, 2019. [Presented at CVPR 2019 Workshop on CV for AR/VR]

[16] Y. Li, X. Chen, Z. Liu, Y. Huang, and L. Wang, "Skeleton-based sign language recognition using graph convolutional networks," in Proc. IEEE ICIP, Abu Dhabi, 2020, pp. 781–785.

[17] J. Forster et al., "RWTH-PHOENIX-Weather: A large vocabulary sign language recognition and translation corpus," in Proc. LREC, Istanbul, 2012, pp. 3785–3789.

[18] D. Li, C. Rodriguez, X. Yu, and H. Li, "WLASL: A large-scale dataset for word-level American Sign Language," in Proc. IEEE WACV, 2020. [Dataset contribution paper]

[19] J. Huang et al., "Video-based Chinese Sign Language recognition using convolutional neural network," in Proc. IEEE SMC, 2015, pp. 2554–2559.

[20] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," Expert Syst. Appl., vol. 164, p. 113794, 2021.

[21] D. Bragg et al., "Sign language recognition, generation, and translation: An interdisciplinary perspective," in Proc. ACM ASSETS, Pittsburgh, 2019, pp. 1–17.

[22] World Health Organization, World Report on Disability. Geneva: WHO Press, 2011.

[23] M. Huenerfauth and V. L. Hanson, "Sign language in the interface: Access for deaf signers," in The Universal Access Handbook, C. Stephanidis, Ed. Boca Raton, FL: CRC Press, 2009, ch. 37.

[24] K. Bantupalli and Y. Xie, "American Sign Language recognition using deep learning and computer vision," in Proc. IEEE Int. Conf. Big Data, Seattle, 2018, pp. 4896–4899.

[25] A. Owoyemi and A. Olabenjo, "Edge-deployed sign language recognition: A feasibility study for mobile Android deployment in sub-Saharan Africa," J. Assistive Technol., vol. 18, no. 2, pp. 89–102, 2024.

[26] Y. Zhang, N. Zhu, and C. Jiang, "EfficientSign: Compact neural architectures for sign language recognition on mobile devices," Pattern Recognit. Lett., vol. 164, pp. 12–19, 2022.

[27] M. Ronchetti, F. Quiroga, C. Estrebou, L. Lanzarini, and A. Rosete, "LSA64: An Argentinian Sign Language dataset," in Proc. XXII CACIC, San Luis, 2016.

[28] A. Sharma and P. Sharma, "Indian Sign Language recognition using deep learning," in Proc. IEEE ICCCNT, Kharagpur, 2020, pp. 1–6.