International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

AI-Enabled Sign Language Recognition For Inclusive Education: A Comparative Analysis Of India And Australia Through Computer Vision And Natural Language Processing

Aastha Thakur[1*] and Dr. A. Marisport[2]

[1] Ph.D. Research Scholar, Gujarat National Law University, Gujarat, India; ORCID ID - 0000-0002-3988-7118, email: aasthathakur613@gmail.com
[2] Assistant Professor of Law, Gujarat National Law University, Gujarat, India. ORCID ID - 0000-0001-6078-6074, email; marisport85@gmail.com
*Corresponding Author: Aastha Thakur

Abstract
Over the last decade there has been rapid convergence of artificial intelligence, computer vision and natural language processing which has presented a significant opportunity to transform communication and educational accessibility for deaf learners over the period of time, yet the sign language technologies remain substantially underdeveloped compared with spoken language AI. In reference to India, Census 2011 has recorded 1.26 million persons with hearing disability, while contemporary research estimates that almost 5 to 6 million users of Indian Sign Language, has reflected a substantial population whose linguistic and educational needs cannot be adequately captured through conventional speech-oriented technologies. There has been a constant technological gap which is evident in the limited availability of high quality, representative datasets that is the INCLUDE data set contains approximately 0.27 million frames across 4287 videos covering 263 ISL signs While recent research has continued to identify data set scarcity dialectal variation and infrastructure constraints as barriers to continuous ISL recognition. However, Australia over the period of time has presented a useful comparative setting, with 16,245 people reporting Auslan use at home in the 2021 Census and 2022 National Deaf Census generating 1215 responses, including 846 Deaf, Deafblind, Deaf-disabled and Hard of Hearing Auslan users; Father there has been an establishment of Auslan corpus comprising approximately 300 hours of videos involving 100 narratives and near native singers. Against all this background has evolved for the study making a comparatively understanding how the AI has enabled sign language recognition through computer vision deep learning and NLP, focusing on data set adequacy, recognition, translation, linguistic diversity, educational integration and human AI interaction. Moreover, this shall argue over the technological accuracy alone is however not sufficient and that the inclusive requires representative datasets, culturally responsive modelling, Privacy safeguards and reproducible evaluation. This however will identify different strategies for developing trustworthy, real time and human cantered AI system mechanism which can be capable of strengthening communication and inclusive education for deaf learners worldwide.

Keywords: Artificial Intelligence, Computer Vision, NLP, Accessibility, Education.

1. Introduction

Development of Artificial Intelligence AI particularly in computer vision deep learning in natural Language processing NLP is creating new possibilities for addressing long standing barriers to communication and education faced by deaf persons. This development is particularly significant against the growing global prevalence of hearing loss the World Health Organization. WHO estimates that more than 1.5 million people approximately one fifth of the world's population currently experience some degree of hearing loss of whom around 430,000,000 require rehabilitation for disabling hearing loss the number is projected to increase to approximately 2.5 billion by 2050 with more than 700 million people expected to require hearing rehabilitation Importantly approximately 95.1 million children between five and 19 years of age currently live with hearing loss these figures establish that accessibility in communication and education is not a marginal technological concern but a substantial global inclusion challenge AI enabled technologies capable of interpreting visual and linguistic communication therefore merit examination is part of the broader development of human cantered and socially responsive AI.

For many Deaf persons, sign language is not simply a substitute for spoken communication; it constitutes a complete linguistic system through which communication, learning, identity and social participation occur. The WHO accordingly identifies specialised education, sign-language instruction and assistive technologies among the interventions capable of mitigating the consequences of hearing loss. The educational dimension is particularly important because barriers to communication during childhood can affect language development, educational participation and later employment and social integration. The significance of sign-language technologies must therefore be understood within the concept of digital inclusion, which requires more than providing access to generic digital devices. It requires digital systems that are capable of recognising and responding to the linguistic modalities actually used by their intended users. In this respect, AI-based sign-language recognition can potentially bridge communication between Deaf learners, teachers and digital learning platforms by converting signed communication into text or other machine-readable outputs (WHO, 2021).

The comparative examination of India and Australia is particularly useful because the two countries represent substantially different linguistic, demographic and technological environments. In India, the INCLUDE (Indian Lexicon Sign Language Recognition) dataset was developed to address the scarcity of publicly available resources for Indian Sign Language (ISL). The dataset contains approximately 0.27 million frames across 4,287 videos representing 263 signs from 15-word categories. The dataset documentation reports that the best-performing model achieved 94.5% accuracy on the 50-sign INCLUDE-50 subset, but only 85.6% accuracy on the larger 263-sign INCLUDE dataset. The difference is analytically important: it demonstrates that AI performance can decline as vocabulary size and linguistic complexity increase and that success in controlled or limited-vocabulary recognition cannot automatically be equated with effective real-world continuous sign-language translation (Sridhar, 2020).
The INCLUDE results demonstrate both the potential and limitations of this technological trajectory. An accuracy of 94.5% for a 50-sign subset indicates that deep-learning systems can achieve strong results under defined conditions, whereas the reduction to 85.6% for 263 signs highlights the increasing difficulty associated with larger vocabularies. Similarly, the existence of approximately 300 hours of Auslan video data demonstrates that naturalistic corpora can provide a foundation for more sophisticated computational modelling, although the existence of data does not itself guarantee that those data are sufficiently annotated, balanced or structured for machine-learning applications (Johnston, 2008). The central technological challenge is consequently shifting from simple classification towards continuous sign-language recognition, contextual interpretation, translation and real-time human-AI interaction.

The technological development of automated sign-language recognition has progressed from rule-based and sensor-dependent systems towards sophisticated approaches based on computer vision, machine learning, deep learning and Natural Language Processing (NLP). Early recognition systems generally relied upon manually engineered visual features and controlled environments, limiting their ability to interpret natural signing. According to the contemporary systems increasingly treat sign-language recognition as a video-understanding problem, where spatial, temporal and linguistic information must be processed together. This distinction is important because sign languages cannot be reduced to isolated hand gestures; meaning may depend upon hand configuration, orientation, movement, facial expression, body posture and the relationship between successive signs. Consequently, modern sign-language AI increasingly combines computer vision with temporal modelling and linguistic processing (Bragg, 2019).
The computer vision forms the first technological layer because AI systems must extract meaningful information from images and video, and the contemporary approaches can identify hand position, hand shape, body posture, facial landmarks and spatial relationships, allowing models to distinguish visually similar signs based on movement and location. However, the large-scale datasets have contributed significantly to this development. The WLASL dataset contains 2,000 sign classes, with standardised subsets of 100, 300, 1,000 and 2,000 words, enabling systematic evaluation of recognition as vocabulary size increases. Such datasets demonstrate that sign-language recognition is moving beyond small, controlled experiments towards more complex and realistic recognition environments (Li, Rodriguez, Yu, & Li, 2020).

Further today the machine learning and deep neural networks have subsequently improved the ability of systems to learn visual representations directly from training data. Convolutional Neural Networks (CNNs) can extract spatial features, while recurrent and transformer-based architectures can model temporal relationships between consecutive frames. This temporal dimension is essential because continuous signing involves sequences rather than independent gestures. Research has consequently begun shifting from isolated-sign classification towards continuous recognition, where systems attempt to identify and interpret unsegmented signing sequences. This development is particularly relevant to educational applications because real classroom communication cannot realistically require users to pause after every individual sign (Adaloglou, et al., 2020).

The NLP becomes important when recognition must progress from identifying signs to generating meaningful text or speech. Sign-to-text translation is not simple word substitution because sign languages possess their own grammatical structures and contextual relationships. Accordingly, an effective system requires computer vision to capture visual information, temporal models to establish sequence and NLP to generate linguistically coherent output. Multimodal AI can further integrate RGB video, hand landmarks, skeletal information, facial expressions and linguistic context, creating opportunities for more natural human-AI interaction (Bragg, 2019). However, the accuracy, explainability and reproducibility are essential for responsible sign-language AI. The accuracy should be assessed through precision, recall, F1-score, confusion matrices and signer-independent testing rather than a single aggregate score.

2. Comparative Analysis Of Ai-Enabled Sign Language Recognition In India And Australia

The comparative development of AI-enabled sign-language recognition in India and Australia reveals two active but substantially different technological and linguistic ecosystems. In India, research on Indian Sign Language (ISL) has increasingly focused on addressing the historical scarcity of machine-readable datasets. The INCLUDE dataset represents an important development, containing approximately 0.27 million frames from 4,287 videos covering 263 signs across 15-word categories. The reported best-performing model achieved 94.5% accuracy on the 50-sign INCLUDE-50 subset and 85.6% on the complete 263-sign dataset. This difference illustrates the increasing difficulty of recognition when vocabulary size expands and demonstrates why results obtained from limited-vocabulary experiments cannot automatically be generalised to continuous, real-world sign-language recognition (Wali, Shariq, & Shoaib, 2023). According to the recent study it further indicates growing academic interest in ISL recognition, translation and deep-learning applications, although dataset diversity and continuous-sign recognition remain significant challenges (Sridhar, 2020).

Today Australia presents a comparatively mature linguistic-resource environment through its Australian Sign Language (Auslan) research infrastructure. The Auslan Corpus contains approximately 300 hours of digital video involving 100 Deaf native and near-native signers, with participants drawn from five Australian locations: Sydney, Melbourne, Brisbane, Adelaide and Perth. The corpus encompasses natural conversations, interviews, storytelling and elicitation activities and includes multiple regional varieties. Its subsequent annotation has incorporated features such as sign tokens, grammatical categories, gaze, palm orientation, clause boundaries and translations. Such naturalistic data are valuable for AI because they provide considerably more linguistic and contextual variation than datasets consisting exclusively of isolated gestures (Johnston, The Auslan Corpus: Current developments and future directions, 2008).

The comparison demonstrates that dataset quantity and dataset usefulness are not identical concepts. INCLUDE was specifically designed as a sign-recognition dataset, whereas the Auslan Corpus originated primarily as a linguistic resource and has progressively acquired greater computational value through annotation (Sridhar, 2020). India's principal challenge remains the expansion of representative ISL datasets, while Australia's challenge includes converting extensive naturalistic linguistic resources into sufficiently annotated and machine-readable datasets. In both jurisdictions, computer vision and deep-learning systems must address variations in signer, speed, background, hand configuration, body movement and vocabulary. Consequently, model evaluation should extend beyond aggregate accuracy towards precision, recall, F1-score, signer-independent testing and cross-context generalisation (Rastgoo, Kiani, & Escalera, 2021).

The development of NLP-based sign-to-text and text-to-sign technologies introduces an additional challenge because recognition is not equivalent to translation. Sign languages possess their own grammatical structures and contextual conventions; therefore, simply identifying a sequence of gestures does not guarantee linguistically meaningful output. Effective systems require the integration of computer vision, temporal modelling and NLP so that visual information can be transformed into contextually appropriate language. The Auslan Corpus, with its translation and grammatical annotations, provides useful resources for such development, while continuing ISL research is expanding the foundations for comparable applications in India (Stoll, Camgoz, Hadfield, & Bowden, 2020).

However, the comparison suggests that AI readiness should not be measured solely through recognition accuracy. India demonstrates considerable research momentum but requires broader and more diverse datasets, while

Australia benefits from established linguistic resources but continues to face challenges concerning annotation, computational transformation and deployment. Both environments must address regional variation, dataset bias, privacy, model generalisability and educational usability. For Deaf education, the real measure of success will be whether AI systems can function reliably in natural classroom environments, support teacher–student communication and provide accessible learning interactions (Wali, Shariq, & Shoaib, 2023). The future therefore lies in developing multimodal, explainable and human-centred AI systems that combine computer vision, deep learning and NLP while respecting the linguistic diversity of Deaf communities.

3. Ai For Inclusive Deaf Education: Ethical, Explainable And Human-Centred Framework

The application of Artificial Intelligence to Deaf education should be understood not merely as an exercise in automating sign recognition but as an opportunity to redesign communication, learning and assessment around accessibility. AI-enabled systems combining computer vision, deep learning and Natural Language Processing can potentially facilitate real-time classroom communication by recognising signed input and converting it into text, captions or other accessible outputs. Such systems could assist communication between Deaf learners and hearing teachers, support participation in classroom discussions and reduce dependence on communication barriers that arise where teachers or peers do not possess adequate sign-language proficiency (Bağcı, 2026). However, AI should function as an assistive layer rather than as a replacement for teachers, interpreters or Deaf linguistic communities. UNESCO emphasises that AI in education should remain human-centred and should protect inclusion, equity, linguistic and cultural diversity and human agency.

Beyond real-time communication, AI can support personalised and adaptive learning for Deaf learners. Machine-learning systems can analyse patterns of interaction and learning performance to provide differentiated educational materials, visual explanations, captions, sign-language resources and adaptive feedback. AI may also facilitate the translation of educational content into accessible formats and assist in generating sign-supported learning materials. Such applications are particularly relevant where educational resources in local sign languages remain limited. UNESCO has recognised the potential of AI to provide personalised support, including for learners with disabilities, while simultaneously warning that unequal access to technology and inadequate safeguards can reinforce existing digital inequalities (Tomlinson, et al., 2026). The objective should therefore be accessibility by design rather than simply adding AI to existing educational platforms after their development.

AI-assisted assessment and feedback present another important area of application. A sufficiently reliable recognition system could potentially interpret signed responses, support formative assessment and provide immediate feedback without requiring every interaction to pass through a human intermediary. Similarly, AI could assist in converting educational materials between text and sign-language representations. Nevertheless, assessment presents a substantially higher risk than simple communication because an erroneous recognition may affect a learner's academic evaluation. Consequently, AI-generated assessment should remain subject to meaningful human review, particularly where the system encounters unfamiliar signing styles, regional variations or low-confidence predictions (Afzaal, et al., 2021). The UNESCO approach to AI in education supports human-centred pedagogical design and ethical validation rather than unrestricted technological substitution.

A central concern is algorithmic bias arising from under-represented sign-language datasets. Sign languages are linguistically diverse, and recognition systems trained predominantly on a limited number of signers, vocabulary categories, geographical regions or recording environments may perform significantly better for represented groups than for users whose signing characteristics are absent from the training data. This problem is particularly important for India and Australia because ISL and Auslan contain regional and contextual variation. Dataset imbalance can consequently transform an apparently accurate AI model into an exclusionary technology when deployed in real classrooms (Prakash & Ram, 2026). Responsible development should therefore involve diverse datasets, signer-independent testing, transparent dataset documentation and meaningful participation of Deaf communities in dataset creation and system evaluation. NIST identifies fairness and harmful-bias management as fundamental characteristics of trustworthy AI.

The use of classroom video and sign-language recordings also creates significant privacy and security concerns. Video data may reveal identifiable faces, bodily characteristics, behavioural patterns and potentially sensitive information about learners. Where such data are collected for model training, informed consent, purpose limitation, secure storage, controlled access and appropriate retention policies become essential. The risk becomes more complex where datasets are reused for purposes beyond the original educational or research objective. NIST's AI Risk Management Framework identifies privacy enhancement, security, transparency, accountability and fairness as interconnected elements of trustworthy AI rather than isolated compliance requirements (Prakash & Ram, Ethical dimensions of AI use in Indian classrooms: Student data privacy vs. personalisation under NEP 2020, 2026). Accordingly, sign-language AI should incorporate privacy considerations at the design stage rather than attempting to address them only after deployment.

Moreover, the explainability is equally important because users and educators should be able to understand the circumstances in which an AI system is confident, uncertain or likely to produce an incorrect recognition. A system reporting a recognition result without communicating its confidence or limitations may create an unjustified perception of reliability. NIST identifies AI systems as trustworthy when they demonstrate characteristics including validity and reliability, safety, security, accountability, transparency, explainability, privacy enhancement and fairness. Applied to Deaf education, this means that AI should provide interpretable outputs, confidence indicators and mechanisms for human correction (Türkmen, 2026). Such feedback can also generate valuable data for subsequent model improvement, provided that corrections are collected ethically and with appropriate consent. On this basis, this study proposes a human-centred AI framework for inclusive sign-language recognition built around seven interconnected stages: community-informed data collection, representative dataset development, multimodal recognition, linguistic interpretation, confidence-based output, human oversight and continuous evaluation. At the first stage, Deaf users, educators and sign-language specialists should participate in defining technological requirements and dataset standards. The second stage should ensure representation across signers, regions, vocabulary and signing conditions. Further the computer vision and deep-learning models should then process hand, body and facial information, while NLP should provide contextual linguistic interpretation. Outputs should incorporate confidence scores and enable teachers or users to correct errors (Altukhi & Pradhan, 2026). Finally, systems should undergo continuous testing for accuracy, fairness, privacy and usability. This framework aligns with the NIST lifecycle approach, which encourages trustworthiness considerations throughout AI design, development, deployment and evaluation.

The future development of this field is likely to be influenced by Generative AI, Large Language Models (LLMs) and Edge AI. Multimodal foundation models may eventually enable systems to connect visual signing with contextual language generation, while LLMs could improve the grammatical and contextual quality of sign-to-text translation (Alam & Khan, 2024). Edge AI could enable recognition to occur directly on smartphones, tablets or classroom devices, reducing latency and potentially limiting the need to transmit sensitive video to external servers. However, generative systems introduce additional risks of hallucination, inaccurate translation, privacy loss and over-reliance on automated outputs. UNESCO therefore advocates a human-centred approach to generative AI in education, including protection of data privacy and ethical and pedagogical validation. The future of inclusive sign-language AI should consequently be measured not simply by computational sophistication but by whether the technology is accurate, explainable, secure, linguistically representative, reproducible and genuinely useful to Deaf learners and educators.

4. Conclusion

The comparative analysis of India and Australia demonstrates that AI-enabled sign-language recognition has considerable potential to strengthen inclusive Deaf education, but its effectiveness depends upon the quality of datasets, technological architecture, linguistic representation and educational context. India has demonstrated significant progress through the development of ISL datasets and deep-learning research, while Australia benefits from comparatively extensive linguistic resources for Auslan. However, both contexts continue to face challenges relating to vocabulary expansion, regional variation, signer diversity, continuous recognition and translation accuracy. The findings indicate that recognition accuracy alone cannot constitute an adequate measure of an inclusive AI system; fairness, explainability, privacy, reproducibility and real-world educational usability must form part of the evaluation framework.

Future AI research should therefore prioritise larger and more representative sign-language datasets, multimodal computer-vision architectures, transformer-based temporal modelling and NLP systems capable of contextual sign-to-text translation. Researchers should adopt transparent reporting standards, signer-independent testing, reproducible experimental protocols and meaningful participation of Deaf communities in system design and evaluation. Privacy-preserving approaches, including secure data governance and Edge AI, should receive greater attention where classroom video is involved. Generative AI and multimodal large language models may further improve contextual translation, but their deployment must remain subject to human oversight and rigorous validation. Ultimately, the next generation of sign-language AI should move from isolated gesture recognition towards reliable, explainable and human-centred linguistic interaction, ensuring that technological innovation contributes meaningfully to educational equality and digital inclusion.

References

1. Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., . . . Daras, P. (2020). A comprehensive study on deep learning-based methods for sign language recognition. IEEE Transactions on Multimedia, 1-43.
2. Afzaal, M., Nouri, J., Zia, A., Papapetrou, P., Fors, U., Wu, Y. L., & Weegar, R. (2021). Explainable AI for data-driven feedback and intelligent action recommendations to support students' self-regulation. Frontiers in Artificial Intelligence, 4, 72.
3. Alam, M. M., & Khan, M. A. (2024). Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication. Artificial Intelligence Review, 57, 188.
4. Altukhi, Z. M., & Pradhan, S. (2026). Integrating explainable AI in education: Challenges and advancements—A systematic review. Pacific Asia Journal of the Association for Information Systems, 1–25.
5. Bağcı, H. (2026). The use of artificial intelligence with d/Deaf and hard of hearing students: A systematic review. American Annals of the Deaf, 390-415.
6. Bragg, D. K. (2019). Sign language recognition, generation, and translation: An interdisciplinary perspective. Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility, 897-931.
7. Johnston, T. (2008). The Auslan Corpus: Current developments and future directions. Sign Language & Linguistics, 191-201.
8. Li, D., Rodriguez, C., Yu, X., & Li, H. (2020). Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 1459–1469.
9. Prakash, D., & Ram, R. K. (2026). Ethical dimensions of AI use in Indian classrooms: Student data privacy vs. personalisation under NEP 2020. Asian Journal of Education and Social Studies, 979–988.
10. Prakash, D., & Ram, R. K. (2026). Ethical dimensions of AI use in Indian classrooms: Student data privacy vs. personalisation under NEP 2020. Asian Journal of Education and Social Studies, 979–988.
11. Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. Expert Systems with Applications, 113794.
12. Sridhar, A. G. (2020). INCLUDE: A large scale dataset for Indian Sign Language recognition. Proceedings of the 28th ACM International Conference on Multimedia, 1366–1375.
13. Stoll, S., Camgoz, N. C., Hadfield, S., & Bowden, R. (2020). Text2Sign: Towards sign language production using neural machine translation and generative adversarial networks. International Journal of Computer Vision, 891–908.
14. Tomlinson, B., Black, R. W., Patterson, D. J., van der Hoek, A., Ferguson, J., & Bietz, M. J. (2026). Students prefer personalized, AI-generated educational videos over non-personalized, human-recorded videos. Scientific Reports, 21804.
15. Türkmen, G. (2026). The review of studies on explainable artificial intelligence in educational research. Journal of Educational Computing Research, 63.
16. Wali, A., Shariq, R., & Shoaib, S. A. (2023). Recent progress in sign language recognition: A Review. Machine Vision and Applications, 34, 37.
17. WHO. (2021). World Report on Hearing. Geneva: WHO.