AI-Based ASL Fingerspelling Recognition Using the Google Kaggle Dataset for Automated Kiosk Transactions

A Literature Review and Project Application

Jamie Pinnington[1,2], Amina Souag[2,3], Hannan Azhar[2,3]

[1]Computer Science, [2]Canterbury Christ Church University, [3]School of Engineering, Technology and Design

Introduction

- According to Ethnologue (2023); Mitchell et al. (2006), there are at least 1 million if not more American Sign Language (ASL) users in the United States alone who are a part of the Deaf and Hard of Hearing (DHH) community.
- Fingerspelling is a critical component of ASL and other sign languages, and is used to spell out words, names, and proper nouns that lack direct signs.
- Current recognition solutions for ASL, don't include fingerspelling, and typically don't work effectively in the real world.
- Legal concerns from the largest and most used ChicagoWild/+ dataset, have made it difficult for commercial application.
- A new dataset released on Kaggle by Google (Manfred Georg et al., 2023) has provided a new opportunity to explore how applying AI could improve, assist, and provide options to the lives of the DHH and the wider community?

View The Paper

Methodology

IEEE Xplore® Digital Library · acm Association for Computing Machinery · ScienceDirect · Google Scholar

Literature Identification

1. Conducted research across four academic databases.
2. Focused on papers published within the last 5 years (2018-2023).
3. Limited to peer-reviewed journal articles, conference papers, and high-quality theses.
4. Search conducted in English only.
5. Search period: November to December 2023.
6. Search terms such as "ASL fingerspelling", "ASL recognition in real-time", "Deep learning for ASL fingerspelling".

Literature Evaluation

1. Abstract → relevance → full reading of selected papers
2. Summarization and analysis of findings in a table
3. Inclusion Relevance to research questions
4. Specific focus on ASL fingerspelling
5. Use of machine learning (ML) models in sign language interpretation
6. Utilization of recognized datasets relevant to ASL recognition
7. Clear methodology, defined objectives, data analysis
8. High citation counts preference
9. Excluded Editorials, opinion pieces, and non-peer-reviewed articles

Results: Figure

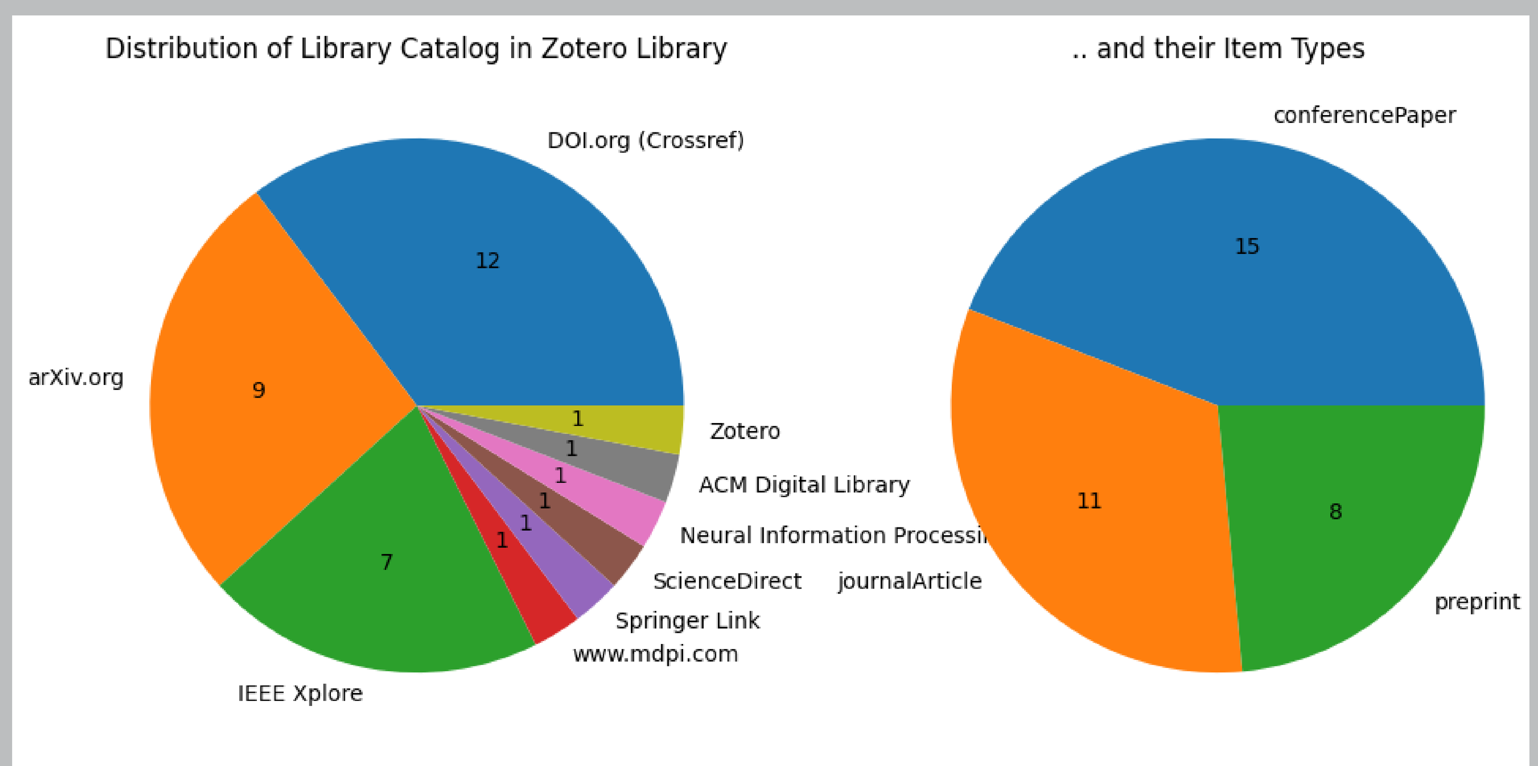


Figure 1: Distribution of Library Catalog and Item Type (2018-2023)

Results: Table

Model	Key Insights and Performance	Citation
RNN, LSTM, Attention	Recognition and translation of ASL glosses; GRR: 86%, GER: 23%. Challenges in real-time recognition.	S Kumar et al. (2018)
CNN, SSD, FCN	High accuracy in vision-based translation; Accuracy: 92.21%. Robustness in ASL recognition.	Abiyev et al. (2020)
Transformers, CTC	State-of-the-art results in ASL recognition; WER, BLEU-4 scores. Translation challenges addressed.	Cihan Camgoz et al. (2020)
ResNet, Bi-LSTM	Recognition using optical flow; Letter accuracy: 57%. Focus on 'wild' conditions and occlusions.	Kabade et al. (2023)
2D/3D-CNN, Bi-LSTM	Superior detection in diverse environments; AP@IoU: 0.495, MSA: 0.386. Fine-grained handshapes analysis.	Shi et al. (2021)
Fine-Grained Visual Attention	Improved recognition with Transformer model; Letter Accuracy: 46.96% (dev). Addressing video data challenges.	Gajurel et al. (2021)

Table 1: Extremely Condensed Summary of ASL Fingerspelling Recognition Models (2018-2023)

Discussion

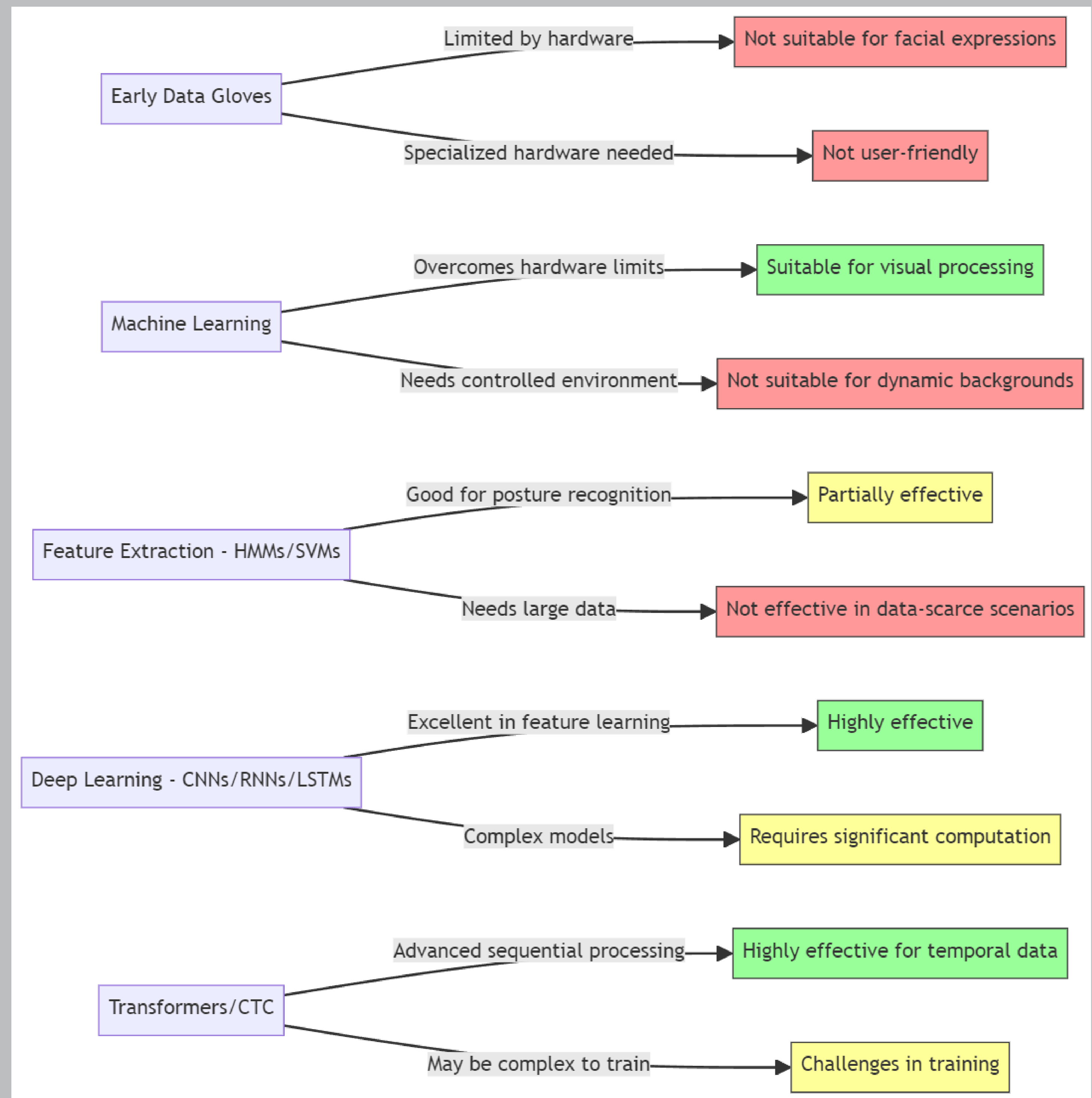


Figure 2: Generalized Overview of Data Analysis of Results (2018-2023)

References

Abiyev, R., Idoko, J.B., Arslan, M., 2020. Reconstruction of Convolutional Neural Network for Sign Language Recognition, in: 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), IEEE, Istanbul, Turkey. pp. 1–5. doi:10.1109/ICECCE49384.2020.9179356.

Cihan Camgoz, N., Koller, O., Hadfield, S., Bowden, R., 2020. Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA. pp. 10020–10030. doi:10.1109/CVPR42600.2020.01004.

Ethnologue, 2023. American Sign Language — Ethnologue Free. URL: https://www.ethnologue.com/language/ase/.

Gajurel, K., Zhong, C., Wang, G., 2021. A Fine-Grained Visual Attention Approach for Fingerspelling Recognition in the Wild, in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Melbourne, Australia. pp. 3266–3271. doi:10.1109/SMC52423.2021.9658982.

Kabade, A.E., Desai, P., C, S., G, S., 2023. American Sign Language Fingerspelling Recognition using Attention Model, in: 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), pp. 1–6. doi:10.1109/I2CT57861.2023.10126277.

Manfred Georg, Mark Sherwood, Phil Culliton, Sam Sepah, Sohier Dane, Thad Starner, Ashley Chow, Glenn Cameron, 2023. Google - american sign language fingerspelling recognition. URL: https://kaggle.com/competitions/asl-fingerspelling.

Mitchell, R.E., Young, T.A., Bachelda, B., Karchmer, M.A., 2006. How Many People Use ASL in the United States?: Why Estimates Need Updating. Sign Language Studies 6, 306–335. URL: https://www.jstor.org/stable/26190621, arXiv:26190621.

S Kumar, S., Wangyal, T., Saboo, V., Srinath, R., 2018. Time Series Neural Networks for Real Time Sign Language Translation, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, Orlando, FL. pp. 243–248. doi:10.1109/ICMLA.2018.00043.

Shi, B., Brentari, D., Shakhnarovich, G., Livescu, K., 2021. Fingerspelling Detection in American Sign Language, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA. pp. 4164–4173. doi:10.1109/CVPR46437.2021.00415.

Technology and Tools Overview

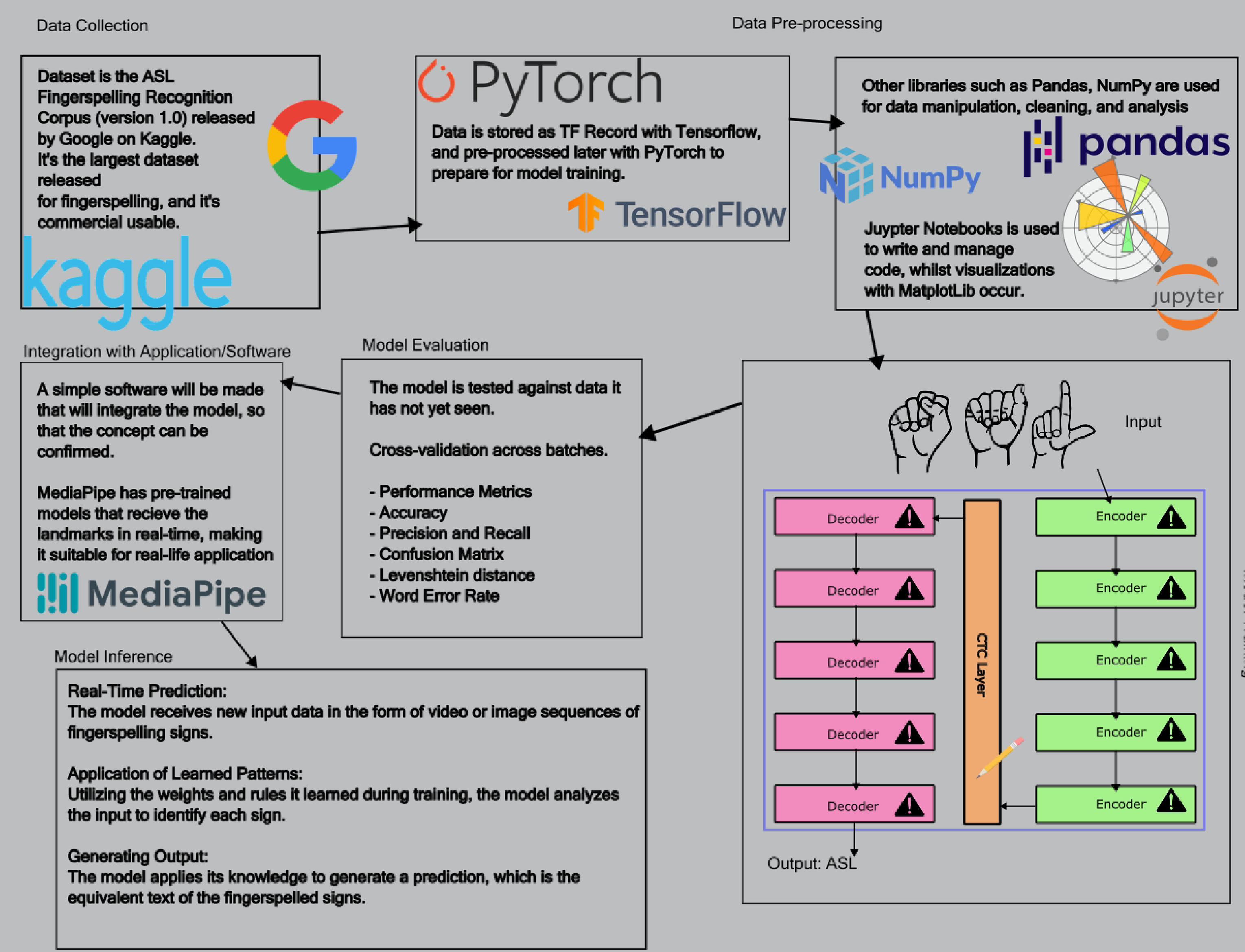
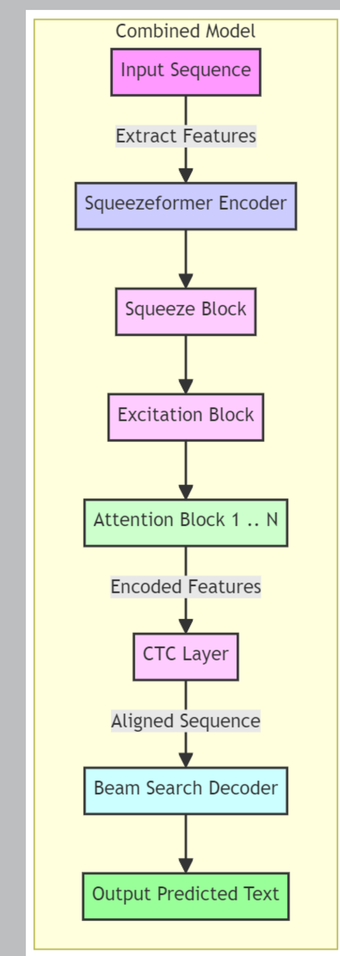


Figure 3: Technology and Tools Overview

Conclusions

1. CNNs, RNNs, LSTMs, and their variants are not sufficient for ASL fingerspelling recognition.
2. Real-world challenges, including rapid signing, varied hand orientations, and diverse environmental conditions, pose substantial difficulties for current models.
3. Techniques such as data augmentation, transfer learning, and the use of pre-trained models are critical in overcoming these obstacles.
4. State-of-the-art Models are variants of Transformer.

Model: The proposed model architecture includes Transformer/Squeezeformer + CTC + Multi-headed Attention + Beam Search.



Contact Information

- Web: www.jamie-pinnington.co.uk
- Email: JP878@canterbury.ac.uk
- Phone: +(44) 07984066009

View Online

Let's Link