DOI: https://doi.org/10.20372/star.V14.i2.03
ISSN: 2226-7522 (Print) and 2305-3372 (Online)
Science, Technology and Arts Research Journal
Sci. Technol. Arts Res. J., April–June 2025, 14(2), 31-38
Journal Homepage: https://journals.wgu.edu.et

Original Research

Static Sign Language Word-Level Detection and Recognition for the Oromo Language

Diriba Negash*, Etana Fikadu & Daditu Dugasa

[1] Department of Computer Science, College of Engineering & Technology, Wollega University, Nekemte, Ethiopia

Abstract

Sign language is a way to communicate ideas and feelings to those who are hard of hearing. In this study, we suggested using AI to help hearing-impaired people communicate better. The study depends on the sign language word level of the Afaan Oromo text. The goal of this paper is to use a deep learning strategy to generate static word-level translations from signed words into equivalent Afaan Oromo texts. Afaan Oromo text is the system's final output, and video frames containing text in signed language serve as the system's input. Our study offers a thorough understanding of how YOLO-v9 functions and outperforms the earlier model. We collected literature, conducted an experiment, and used video data. Pre-processing tasks such as frame extraction, resizing, labeling, and data splitting using Roboflow are carried out in order to train our model. The system achieved a precision of 88.8%, a recall of 91.3%, an mAP of 92.7% at 0.5 IoU, and a score of 75.2% at 0.5:0.95 IoU. In general, our model is usable for our community, who can read Afaan Oromo texts, and the visually impaired to recognize Afaan Oromo, because many people cannot hear and understand the signs.

Article Information

Article History:
Received: 21-11-2024
Revised: 26-03-2025
Accepted: 30-06-2025

Keywords:
Afaan Oromo, signed word level, AI, deep learning, and YOLOv9

*Corresponding Author:

Diriba Negash

E-mail: diribanegash@yahoo.com

INTRODUCTION

Communication is crucial in today's world because it enables both us and others to comprehend information more quickly and properly (Núñez et al., 2023). However, deaf and dumb people find it extremely difficult to interact with regular people (Imran et al., 2024). They convey their message using sign language, which is a combination of gestures, orientations, movements of hands, arms, or body, and facial expressions, but not every normal person can understand that language (Bhat et al., 2022; Farooq et al., 2021). Sign language can be recognized using either of the two approaches. The first one is through the hardware-based system. In this approach, the user is required to wear gloves. The second approach is the vision-based approach, where the gesture is recognized using the concept of computer vision (Núñez et al., 2023). In Ethiopia, Oromo is the most important national language, which is used in trade (Dinsa, 2007; Das, Abebe, et al., 2024), education (Fikadu, 2020), health (Fikadu et al., 2025; Meng et al., 2020), local government, and the mass media (Dinsa et al., 2024b). The Afaan Oromo is one of the official languages of Ethiopia. It is native to the Oromia region of Ethiopia.

It has characters, like vowels, double consonants, and consonant phonemes, i.e., sounds that make a difference in word meaning (Dinsa et al., 2024a). The combinations of those are the basis of the language, giving us phrases, clauses, and sentences (Fikadu et al., 2019). However, 40% of Ethiopians speak Afaan Oromo, unless between 335,000 and 1 million Deaf individuals are in the country (Dulume et al., 2022; Garoma, 2024). Sign language, a sophisticated, visually focused language that is as expressive and nuanced as any spoken language, is central to the culture of deaf people (Bhat et al., 2022). Sign languages use body postures, facial emotions, hand forms, and movements to transmit messages (Pathak et al., 2022). Every nation, or even a region within a nation, has a unique sign language (Weldemikael, 2020). In the absence of hearing people, sign language is an effective tool that offers a channel for artistic expression, education, and cultural preservation in addition to communication (Negash et al., 2023; Fikadu et al., 2025). The objective of our experiment is to help those people who are unable to hear and unable to speak to make conversation with other normal people (Abdalla et al., 2020).

Statement of the problem

In Ethiopia, when hard-of-hearing people talk in sign language, many people can't understand them. Even if hearing-impaired people have the problem of writing and reading Afaan Oromo text, by using new technology, we can minimize all these problems.

Research Question

Our research question is, "How can we design a system that supports equal communication and participation? So, the objective of this work is to use the AI method to improve communication barriers between deaf and hearing people in those communities that can speak and write the Afaan Oromo. We have proposed a model that can recognize and detect static sign gestures by converting them into corresponding words.

MATERIALS AND METHODS

The first step was collecting the videos of data for 54 words in a way that helped us extract all the necessary information for Oromo signed language. The second pre-processing step is using Roboflow tools to label, augment, and export the data directly in the format of YOLOv9. For extraction, the frame of the video using OpenCV-Python has the option for creating image sequences from the video, and then, after extracting the frame, labeling frames is conducted by using LabelImg annotation tools to label manually. Once a box is created, the data regarding its position is stored in a .txt file. By using Roboflow, the data was separated into 70% for the training set, 20% for validation, and 10% for test images. The dataset consists of images in .jpg format and .txt format. Box coordinates must be normalized between 0 and 1. These *.txt files include annotations of bounding boxes of signs in the YOLO format.

Modern deep learning techniques concentrate on creating the best possible goal functions to ensure that the model's predictions are as close to the ground truth as possible (Chen et al., 2024). As shown in Figure 1, a suitable architecture that can make it easier to gather sufficient data for prediction must be created in the meantime (Kodandaram et al., 2021). In several domains, including computer vision, language processing, and speech recognition, deep learning-based models have proven to perform significantly better than previous artificial intelligence systems (Liau, 2021). Strong competitiveness has been demonstrated by the YOLOv9 (Imran et al., 2024), which was created by integrating PGI and GELAN. When compared to YOLOv8, the deep model's superior design enables it to cut the number of parameters by 49% and the number of calculations by 43%. On the MS COCO dataset, there is still a 0.6% AP improvement, nevertheless. As shown in Figure 1, this model incorporates the special integrated structures to assess and identify spoken language translation systems in signed languages.

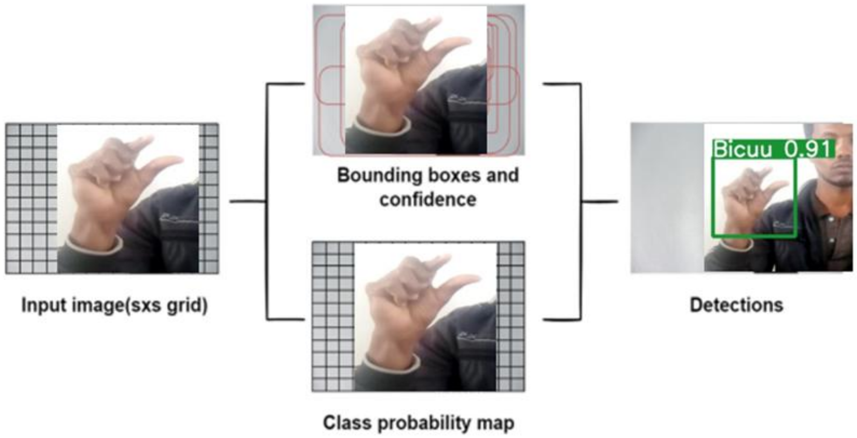

Figure 1. *Architectural diagram of the proposed system*

YOLOv9 architecture

As shown in Figure 1. YOLOv9 optimizes feature extraction and prediction by introducing significant improvements to its head, neck, and backbone components (Wang et al., 2024; Imran et al., 2024). The model strikes the ideal balance between detection speed and accuracy by using the most recent developments in activation functions, normalization approaches, and layer design (Imran et al., 2024).

Training the Model

From the total of 4066 images, 2.8k images are for training purposes, and in the validation set, 816 images are for training. For testing the model, two types of testing were used, namely, testing of image data and video data. In the image testing, we have used 408 images from this image; several are detected. At the end of the testing, a comparison of the detection results with the ground truth will be carried out, creating a confusion matrix to obtain accuracy, precision, and recall. In this, the mean average precision is used to evaluate the sign detection task. In video testing, detection will be performed on each frame of the video data (Liang et al., 2023).

RESULTS AND DISCUSSIONS

Results

In this study, we assemble the YOLOv9 model to recognize and detect signed Afaan Oromo from our dataset. In Figure 2, to evaluate the effect of the proposed model, we used recall, precision, and mAP to quantitatively analyze our model.

gelan-c summary: 467 layers, 25453365 parameters, 0 gradients, 102.7 GFLOPs

Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 26/26 [00:22<00:00, 1.16it/s]
all	816	815	0.89	0.912	0.928	0.752
Abbaa	816	16	0.978	1	0.995	0.681
Abbaa Manaa	816	17	0.936	0.856	0.935	0.589
Adurree	816	12	0.971	1	0.995	0.871
Afur	816	5	0.923	1	0.995	0.874
Akkam	816	23	0.916	0.949	0.961	0.767
Amma	816	23	0.985	1	0.995	0.745
Bicuu	816	21	0.981	1	0.995	0.841
Bilbila	816	23	0.983	1	0.995	0.723
Bishaan	816	21	0.978	1	0.995	0.954
Booda	816	6	0.943	1	0.995	0.717
Boor	816	12	0.774	0.833	0.942	0.666
Dallansuu	816	22	0.98	1	0.995	0.702
Dansaa	816	17	0.987	1	0.995	0.816
Dheeraa	816	4	0.882	1	0.995	0.809
Dhukkubsataa	816	17	0.759	0.471	0.729	0.607
Dibaabee	816	26	0.97	1	0.995	0.951

Figure 2. *Performance of the proposed model*

Inference

Testing is conducted with images, webcams, and videos. For inference, we invoke those weights along with a conf. Specifying model confidence and an inference source. The source can accept a directory of images, individual images, video files, and also a device's webcam port (Negash et al., 2023; Imran et al., 2024). After the detection is complete, the predicted bounding boxes that cover the objects will be drawn into the image. They were saved in the same folder containing results from the training phase.

Confusion Metric

According to Chen et al. (2024), the mean average precision (mAP) and average precision (AP) metrics are used to assess the effectiveness of sign language detection and localization (Padilla et al., 2021). Average accuracy (AP) values are computed as the mean of recall values between 0 and 1. As presented in Figure 3, the confusion matrix, intersection over union (IoU), recall, and precision measurements serve as the foundation for the AP formula.

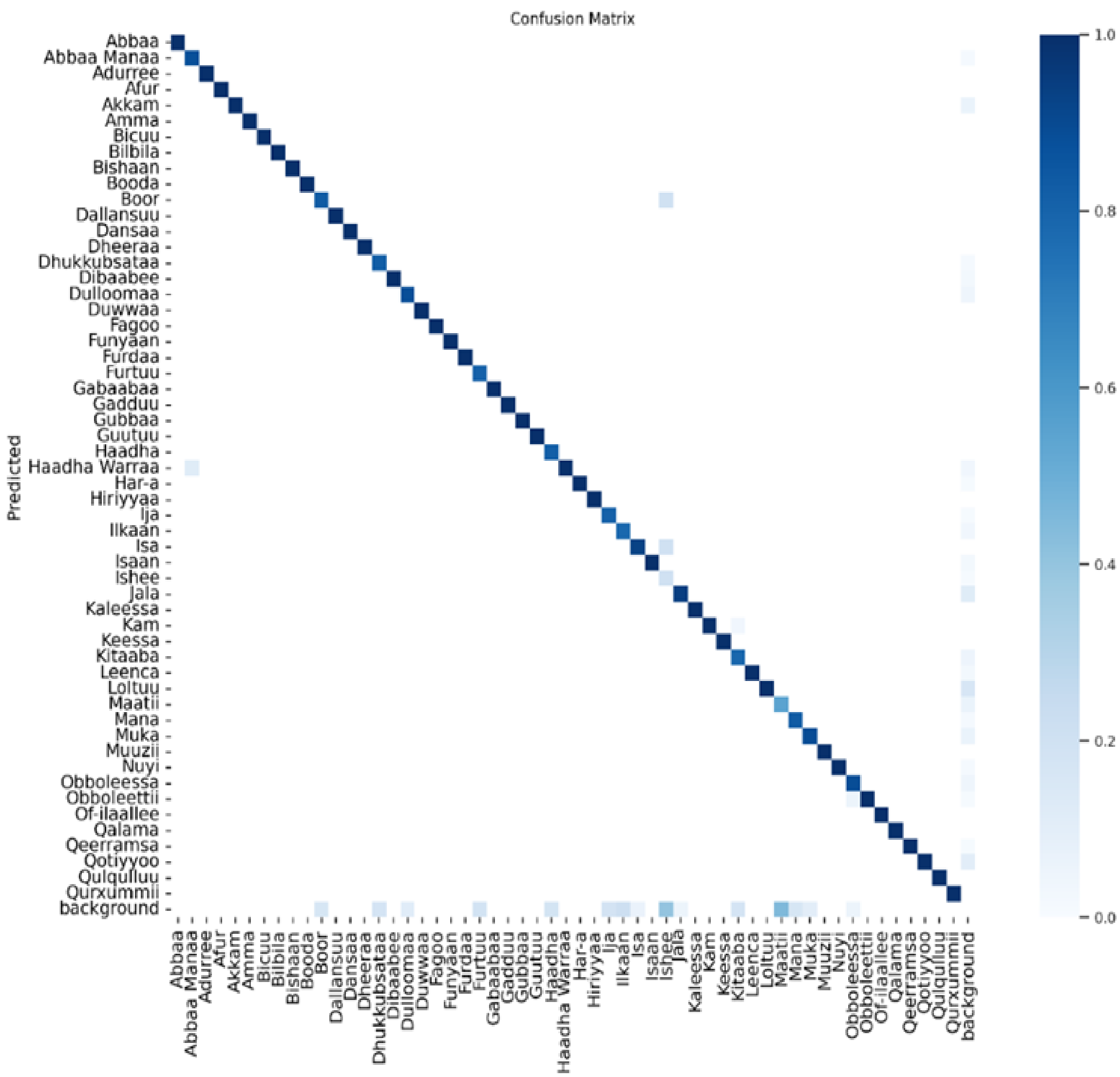

Figure 3. *Confusion matrix*

Discussions

After thorough analysis, the static gestures were trained on YOLOv9, as presented in Fig. 3, which implies that as training went on, the model learned and enhanced its capacity to identify and categorize items. The researcher (Olkeba & Worku, 2021) recognized Amharic characters using Faster R-CNN and SSD. The researcher (Bedaso & Hussein, 2024) accomplished a result in recognizing Afaan Oromo word signs by using AlexNet and - - GoogleNet. But, in our study, the proposed model used YOLOv9, which can localize and classify depending on the region of interest, then detect and recognize Afaan Oromo word level.

A well-performing model is shown by the high precision, rising recall, and rising mAP values as presented in Figure 4. Nonetheless, the model would have profited from more training time.

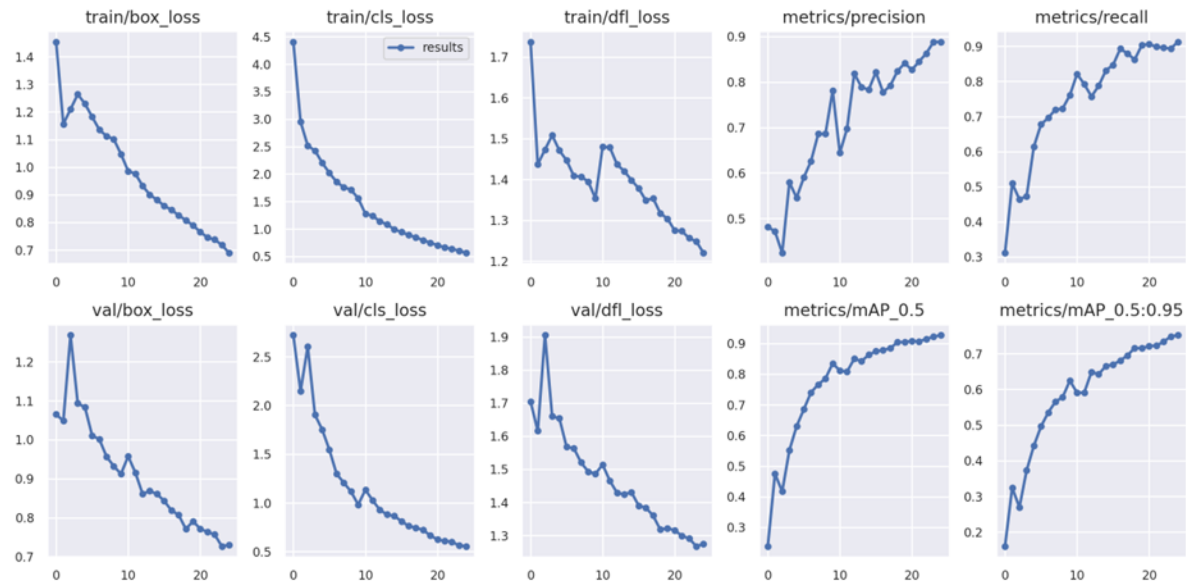

Figure 4. *Visualization of validation rate*

The current study aims to enhance the effectiveness of Afaan Oromo word-level sign language translation by employing a deep learning model (Imran et al., 2024). As shown in Fig. 2, the model architecture to be trained with YOLOv9-c and the training process for 50 epochs were completed with 92% accuracy. The application of deep learning techniques in sign language translation is still unexplored and nascent (Abdalla et al., 2020).

CONCLUSIONS

In this work, we adapted Afaan Oromo's signed word level to texts and highlighted the communication gap between deaf and hearing people. Translation studies in sign language are still being conducted. The basic idea of sign language communication for those with speech impairments is described, along with the challenges they encounter when there is no translator available. Thus, the project's objective is to create a sign language translator that will make it simple for them to communicate with non-signers.

Sign language translation makes use of CNN and YOLOv9 ideas for face expression identification and hand gesture categorization. The model is also flawed in several ways. As shown in Fig. 4, the new sign gesture translation technology can identify gestures from videos in real time. All of the selected words were accurately predicted. There is general agreement that our model can and must be employed in designing a step toward expanding the availability of educational materials for our deaf community. This paper examined the application of artificial intelligence to facilitate spoken language and Afaan Oromo sign language communication. The findings of the study demonstrate that artificial intelligence has progressed to the point where image recognition technology can now translate sign language into written text. Based on the proof of concept that we realized, it can be concluded that an application can be brought to the market that converts Afaan Oromo signed language into spoken or written text. The application then looks through the camera of a laptop or smartphone at the gestures that the gesture-maker performs and converts these into words using image recognition technology.

Recommendations

Based on the results and findings of this study, the researcher recommends that other researchers work on preparing more datasets for this specific language, two-way communication to translate signs to text and vice versa, and sentence-level translation as future work.

CRediT authorship contribution statement

Diriba Negash: Writing—Original Draft, Writing—Review & Editing, **Etana Fikadu:** Conceptualization, Investigation, Resources, Data Curation, Visualization, **Daditu Dugasa:** Formal analysis, Methodology, Validation

Declaration of competing interests

The authors affirm that there is no conflict of interest.

Ethical approval

The authors declare that no human participants, their data, or biological material were used in this study.

Data availability statement

Adequate data are available and will be presented upon request.

Acknowledgment

The author would like to thank the Department of Computer Science, Wollega University, for providing the necessary facilities.

REFERENCES

Abdalla, S., Hassan, H., Fahad, A. A., Hassan, H. J., & Abdullah, S. H. (2020). Deep Learning-based Deaf & Mute Gesture Translation System. *Article in International Journal of Science and Research*, *5*, 3–8. https://doi.org/10.21275/SR20503031800

Bedaso, M., & Hussein, A. (2024). *Developing Sign Language Recognition Model for Afaan Oromoo Words Using a Deep Learning Techniques*. 1–23.https://doi.org/10.21203 /rs.3.rs-4218445/v1

Bhat, A., Yadav, V., Dargan, V., & Yash. (2022). Sign Language to Text Conversion using Deep Learning. *2022 3rd International Conference for Emerging Technology, INCET 2022*, 4036–4044. https://doi.org/10.1109/INCET54531.2022.9824885

Chen, W., Luo, J., Zhang, F., & Tian, Z. (2024). A review of object detection: Datasets, performance evaluation, architecture, applications and current trends. In *Multimedia Tools and Applications* (Vol. 83, Issue 24). Springer US. https://doi.org/10.1007/s11042-023-17949-4

Dinsa, E. (2007). *Innovative Factors and Challenges in Integrating Technology in Education for Global Development*. 1028–1035. https://doi.org/10.5281/zenodo.6913186

Dinsa, E. F., Das, M., & Abebe, T. U. (2024a). A topic modeling approach for analyzing and categorizing electronic healthcare documents in Afaan Oromo without label information. *Scientific Reports*, *14*(1), 1–14. https://doi.org/10.1038/s41598-024-83743-3

Dinsa, E. F., Das, M., & Abebe, T. U. (2024b). OPEN AI - based disease category prediction model using symptoms from low - resource Ethiopian language : Afaan Oromo text. *Scientific Reports*, *May*. https://doi.org/10.1038/s41598-024-62278-7

Dinsa, E. F., Das, M., Abebe, T. U., & Ramaswamy, K. (2024). Automatic categorization of medical documents in Afaan Oromo using ensemble machine learning techniques. *Discover Applied Sciences*, *6*(11). https://doi.org/10.1007/s42452-024-06307-0

Dulume, W., Justice, O., & Profesionals, S. (2022). Assessing Key Factors in the Making Afan Oromo One of the Federal Working Languages. *International Affairs and Global Strategy*, *October*. https://doi.org/10.7176/iags/93-01

Farooq, U., Rahim, M. S. M., Sabir, N., Hussain, A., & Abid, A. (2021). Advances in machine translation for sign language: approaches, limitations, and challenges. *Neural Computing and Applications*, *33*(21), 14357–14399. https://doi.org/10.1007/s00521-021-06079-3

Fikadu, D. E., & Babu P.R. (2019). Application of Data Mining Classification Algorithms for Afaan Oromo Media Text News Categorization. *International Journal of Computer Trends & Technology*, *67*(7), 73–79. https://doi.org/10.14445/22312803/ijctt-v67i7p112

Fikadu, E. (2020). Blended- Learning Approach for Ethiopian Education System : In Case of Second Generation University. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *6*(3), 947–954. https://doi.org/10.32628/cseit2063206

Fikadu, E., Das, M., Urgessa, T., & Ramaswamy, K. (2025). A supervised learning approach for recommending medical specialists in the healthcare sector for the Afaan Oromo context. *Discover Computing*, *28*(1). https://doi.org/10.1007/s10791-025-09556-8

Fikadu, E., Urgessa, T., & Das, M. (2025). MLP-SVM: a hybrid approach for improving the performance of the classification model for health-related documents from social media using multi-layer perceptron and support vector machine. *Discover Applied Sciences*, *7*(4). https://doi.org/10.1007/s42452-025-06851-3

Garoma, E. T. (2024). Demonstratives in Afaan Oromoo. *Cogent Arts and Humanities*, *11*(1). https://doi.org/10.1080/23311983.2023.2297494

Imran, A., Hulikal, M. S., & Gardi, H. A. A. (2024). *Real Time American Sign Language Detection Using Yolo-v9*. http://arxiv.org/abs/2407.17950

Kodandaram, S. R., Kumar, N. P., & Gl, S. (2021). *Sign Language Recognition*. *July*. https://doi.org/10.13140/RG.2.2.29061.47845

Liang, Z., Li, H., & Chai, J. (2023). Sign Language Translation: A Survey of Approaches and Techniques. *Electronics (Switzerland)*, *12*(12). https://doi.org/10.3390/electronics12122678

Liau, Y. Y. (2021). *Status Recognition Using Pre-Trained YOLOv5 for Sustainable Human-Robot Collaboration (HRC) System in Mold Assembly*.https://doi.org/10.3390/su132112044

Meng, Y., Speier, W., Shufelt, C., Joung, S.E., Van Eyk, J., Bairey Merz, C. N., Lopez, M., Spiegel, B., & Arnold, C. W. (2020). A Machine Learning Approach to Classifying Self-Reported Health Status in a Cohort of Patients with Heart Disease Using Activity Tracker Data. *IEEE Journal of Biomedical and Health Informatics*, *24*(3), 878–884. https://doi.org/10.1109/JBHI.2019.2922178

Negash, T.D., Fikadu, D.E., & Fikadu, K. H. (2023). Signed Language Translation into Afaan Oromo Text Using Deep-Learning Approach. *American Journal of Artificial Intelligence*, *7*(2), 40–51. https://doi.org/10.11648/j.ajai.20230702.12

Núñez, M. A., Perez, V.O., & Labaka, G. (2023). A survey on Sign Language machine translation. *Expert Systems with Applications*, *213*, 118993. https://doi.org/10.1016/j.eswa.2022.118993

Olkeba, I. F., & Worku, H. S. (2021). Developing Amharic Sign Language Recognition Model for Amharic Characters Using Deep Learning Approach. *Research Square*, 1–18. https://doi.org/10.21203/rs.3.rs-236824/v1

Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & Da Silva, E. A. B. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*

(Switzerland), *10*(3), 1–28. https://doi.org/10.3390/electronics10030279

Pathak, A., Kumar, P.A., Chugh, G., & Ijmtst, E. (2022). Real Time Sign Language Detection. *International Journal for Modern Trends in Science and Technology*, *8*, 32–37. https://doi.org/10.46501/IJMTST08 01 006

Wang, C.Y., Yeh, I.H., & Liao, H.Y. M. (2024). *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*. https://doi.org/10.1007/978-3-031-72751-1_1

Weldemikael, F. (2020). *Ethiopian Deaf Associations'Perception and Satisfaction Towards Sign Language Interpretation Service of Ethiopian*. https://doi.org/10.1353/sls.2018.0037