JCSDA, Vol. 1, No. 2, 76–95
DOI: 10.69660/jcsda.01022405
ISSN 2959-6912

A Multi-modal Fusion Technique to Combine Manual and Non-Manual Cues for Amharic Sign Language Recognition: A Systematic Literature Review

Isayas Feyera

Dep't of Software Engineering,
Addis Ababa Science and Technology University,
Addis Ababa, Ethiopia
Coresponding author: isayas.feyera@aastu.edu.et

Solomon Teferra

Addis Ababa University, Ethiopia

Asrat Mulatu Beyene

Artificial Intelligence and Robotics CoE,
Dep't of Electrical and Computer Engineering,
Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Amharic Sign Language (ASL) is a vital form of communication for the hearing-impaired community in Ethiopia. Recognizing and understanding ASL is crucial for facilitating communication and accessibility for Amharic-speaking hearing-impaired individuals. ASL relies on manual gestures, including hand shapes and movements, and non-manual cues such as facial expressions and body postures. The objective of this review is to investigate the methodologies employed to combine manual and non-manual cues in ASL recognition systems. In our review, we have considered various inclusion and exclusion criteria to select relevant research papers. After primary data selection, 46 papers which focus on sign language are included in our analysis. We also employed a data extraction form to collect and gather information from these selected papers systematically. Based on the review, combining manual and non-manual cues enhances the accuracy and robustness of sign language recognition systems. These techniques leverage computer vision and machine learning approaches to interpret manual gestures, while also capturing the nuanced information conveyed through facial expressions and body language. Improving hand gesture recognition involves the finding of key points or poses. Despite the advancements in ASL recognition, this review underscores a significant challenge—the lack of available resources, reputable publications, annotated data, and annotating tools specific to Amharic Sign Language are scarce. This shortage hampers the development and evaluation of ASL recognition systems, hindering progress in this field.

Keywords*: Hybrid Cryptographic Algorithm, Cloud Data Security, Security Performance Evaluation, Asymmetric Algorithm, Symmetric Algorithm, Data Integrity Verification.*

1. Introduction

Fusion techniques refer to methods and approaches used to integrate and combine information from multiple modalities or sources of data. In the context of Amharic sign language recognition, multi-modal fusion involves combining man-

ual and non-manual cues, such as hand shapes, movements, facial expressions, and body gestures, to improve the accuracy and robustness of the recognition system. Sign languages are natural languages that vary from signer to signer. Manual and non-manual sign language recognition refers to the process of interpreting both the manual (hand-based) and non-manual (facial expressions, body movements, and other non-hand gestures) aspects of Amharic sign language. In Amharic sign language, manual cues primarily include hand shapes, and movements, while non-manual cues play an important role in, emotions, and nuances of the language [1]. In the world, more than four hundred sixty-six million (466 million) people have lost their hearing as reported by the World Health Organization [2]. It accounts for more than 5% of the world's population, with children accounting for 5% (34 million). There will be approximately 630 million people with disabilities and hearing loss by 2030 unless action is taken. This number will grow to 900 million by 2050 G.C [3]. A genetic disorder or cause could cause hearing loss. According to the report, 1.1 billion people aged 12 to 35 will be exposed to hearing loss as a result of excessive noise.

Humans, by nature, have a voice for communication. However, some people may have difficulty of communicating due to hearing loss or becoming deaf in various circumstances [4]. Nature, accidents, or advances in age are all possible causes. Hearing-impaired people use their eyes to recognize Sign Language. According to the paper [5] during communication, there is about 55% of useful information from facial expressions, 38% from sound, and 5% from conveying language. Sign Language is a communication system in which people with hearing loss and those with normal hearing interpret communication visually to exchange information. Approximately 0.4% of Ethiopians are deaf, and nearly 50% are illiterate[5]. This figure indicates a dearth of research into hearing-impaired people in Ethiopia.

2. Statement of the problem

There is a lack of annotated data for Amharic sign languages, making it difficult to train machine-learning algorithms for Amharic language sign recognition. Recognizing signs also requires processing large amounts of visual information, which can be computationally complex and demanding.

Despite these challenges, researchers are working towards developing manual and non-manual sign language recognition systems, using various technologies/algorithms, such as deep learning, computer vision, and machine learning. The goal is to develop systems that can recognize signs performed by different signers and in different signing styles[orientation, location], making sign language more accessible to a wider audience.

in our country Ethiopia, Amharic Sign Language (ASL) serves as a vital means of communication for the hearing-impaired community. However, detecting, recognizing, and interpreting ASL gestures accurately presents a huge challenge due

to the need to effectively combine both manual and non-manual cues. The problem at hand is the limited understanding and development of multimodal fusion techniques that can integrate manual and non-manual cues based on multimodal fusion for ASL recognition. Signing is a continuous, dynamic process, and even the same sign can be performed in different ways [Position, Orientation, Location] by different signers, making sign recognition challenging.

2.1. *Hand Orientation*

Hand orientation is the position of the palm having various positions. When the palm orientation changes, the meaning also changes, The orientation of the hand indicates that when the palm faces in, out, horizontally, left, up, or down [6].

Fig. 1. Different hand orientation for the Amharic character[6]

2.2. *Hand location*

During communication with sign language, the movement of the hand can be done via different directions to represent messages, These are locations around the forehead, chest, left to right, and right to left[6].

Fig. 2. Hand locations for representing a letter derived from [6]

3. Objective

The objective of this literature review is broken down into two parts mainly general and specific objectives.

3.1. *General Objective*

The general objective of this systematic literature review is to investigate the state of research on the recognition of Amharic Sign Language (ASL) through a multi-modal fusion combination of manual and non-manual cues.

3.2. *Specific Objectives*

- To find methods to extract relevant features from the sign language gestures, such as hand shapes, hand movements, and facial expressions.
- To identify the existing research studies focused on manual recognition of Amharic Sign Language.
- To explore the research on non-manual recognition of Amharic Sign Language.
- To find the challenges of multimodal fusion in ASL recognition.
- To assess the current state-of-the-art and future directions for ASL recognition.

4. Research Questions

(1) How can hand gesture recognition be improved?
(2) Which features provide better accurate recognition of sign language in Amharic Sign Language?
(3) What are the most widely used deep learning algorithms for manual and non-manual sign language recognition?

5. Related Works

Yigremachew Eshetu's [2] work discusses a real-time Ethiopian Sign Language translation to audio converter using a hybrid of vision-based and sensor-based area units to collect hand configurations and knowledge of the accompanying meaning of gestures. The paper uses the pointed technique for the following purposes, which will be explained separately. The first is a vision-based solution, in which they employ a single camera to track the user's hands to recognize them. Sign Language in Ethiopia on TensorFlow, the system using a machine learning neural network dubbed Single Shot Multi-Box Detector (SSD). This network assisted them in detecting hand motions. The essential stage for this vision-based sensor is face detection through the haar-cascade [7] [3] to justify whether there are who are going to sign, hand detection, and overlap detection. This overlap detection means that every time the system takes a real-time video, it recognizes and analyzes the footage to see if there is an overlap of hands and faces in Ethiopian Sign Language meaning. For example, if someone places his hand on his chin, this means 'Mother,' and if someone places his/her hands on his/her forehead, this means 'Father,' as seen in the photographs below. This means that, according to the SSD method, the

hands and forehead are overlapping, hence the meaning of Ethiopian Sign Language relies on where the user's hands are situated. [8]

Fig. 3. ASL: 'āmeseginalihu' (Meaning Thankyou)

Amharic Sign Language (ASL) serves as a vital means of communication for the deaf community in Ethiopia. In recent years, there has been an increasing interest in developing ASL recognition, leveraging both manual and non-manual cues. This systematic literature review aims to explore the existing research in this domain, providing an overview of the advancements, challenges, and future directions for multimodal fusion to combine manual and non-manual recognition of ASL. The

Fig. 4. ASL: 'āmeseginalihu' (Meaning Thankyou)

study [9] addresses the existing gap in the literature by conducting the first identifiable academic Systematic Literature Review (SLR) and proposing a classification scheme for sign language recognition systems. The review covers a period from 2007 to 2017 and encompasses 396 research articles, which were analyzed for their relevance to sign language recognition systems. Following the review process, 117 articles were selected, reviewed, and classified. The classification scheme involved categorizing the papers based on 25 sign languages and comparing them across six dimensions, including data acquisition techniques, static/dynamic signs, signing mode, single/double-handed signs, classification technique, and recognition rate. The findings of the review reveal that the research in sign language recognition focuses on static, isolated, and single-handed signs captured using cameras. The study aims to provide a road map for future research and contribute to the accumulation and creation of knowledge in the field of sign language recognition.

[10] utilizes simple low-level processes operating on images to build realistic representations, which are then fed into intermediate-level processes to form sign hypotheses. At the intermediate level, the researchers construct representations for both the manual and non-manual aspects of American Sign Language (ASL), including hand movements, facial expressions, and head nods. The integration of manual and non-manual information is performed sequentially, with the non-manual information refining the set of manual information-based hypotheses. The results of the study demonstrate that using image-based manual information alone achieves a correct detection rate of approximately 88%. However, when facial information is incorporated, the accuracy increases to 92%, highlighting the valuable contribution of facial cues in ASL recognition. Additionally, the researchers were able to successfully detect instances of "negation" in sentences with a 90% accuracy using only 2D head motion information.

The objective of the research [11] is to conduct a Systematic Literature Review (SLR) to gather and synthesize existing knowledge and experiences related to Vision-Based Recognition (VBR) techniques. The review specifically emphasizes hand gesture recognition (HGR) techniques and enabling technologies. Through a careful selection process, 100 relevant studies were chosen and subsequently analyzed to identify the current state-of-the-art in vision-based gesture recognition. The researchers aim to provide valuable insights and highlight key findings in the field of vision-based gesture recognition, with a specific focus on HGR techniques. This systematic review offers a comprehensive overview of the existing research, enabling a better understanding of the advancements and challenges in this area of HCI technology. The author of [12] has used to combine both manual and non-manual features of sign language to extract facial expressions. To extract the feature of facial expression they have used an active appearance model in addition to a supporting vector machine for the recognition of non-manual features. They have the

highest recognition rate of 84%. The main aim of the paper was to combine the performance of manual and non-manual features of sign language they have created two different conditions: non-manual only and manual and non-manual sign language features combined.

6. Hearing-impaired People

Hearing-impaired people communicate using Sign Language and nonverbal communication. These folks primarily communicate with hearing-impaired and normal-hearing people via Sign Language. These nonverbal communications include gesture movement, head tilt, finger spelling, shoulder shrugging, rising eyebrow, and other nonverbal communication. In everyday life, people communicate by shrugging off their shoulders or using other non-manual actions. Some non-manual Sign Language has a challenging scenario to consider as Sign Language since Sign Language is an organized method of communication in which every word or alphabet is sorted and assigned a gesture. Hand gestures and eyebrow raises, whether known or unknown, can be utilized to communicate with another person in some instances [13].

7. Sign Language

The presence of American and Nordic missionaries who built hearing-impaired schools in Ethiopia in the 1960s resulted in the emergence of Sign Language. [14]. Despite popular belief that Sign Language is not a natural language, it is a natural language with its own set of laws. As a result, Sign Language, like any other language, can be utilized by hearing-impaired, stupid, or normal-hearing people. The significance of Sign Language stems from the fact that it was used by early humans before the emergence of vocal language and computers today.

Before learning the mother tongue, a kid uses or communicates gesture communication when he or she requires food, warmth, or comfort. As a result, this Sign Language can be learned with a gesture from the start. Sign Language is a vision-based language that is typically used by the deaf. The current scenario reveals Sign Language is often utilized or used worldwide such as in international sports to read the player's Sign Language if it is used inappropriately, and it can also be used in religious practice, for traffic signs on the road, and in a variety of other ways. The structure of Sign Language varies by country: American Sign Language, Indian Sign Language, Arabic Sign Language, British Sign Language, and other languages are available [15].

8. Ethiopian Sign Language (EthSL)

According to [17], EthSL is derived from American Sign Language, but it also has some influence from Nordic Sign Languages like Finnish Sign Language. This implies that there is some small and specific local Sign Language used by some hearing-impaired Ethiopians and then incorporated into Ethiopian Sign Language. The local

Fig. 5. Facial Expression[16]

Sign Language, which was first taught in Ethiopia by American missionaries and was based on American Sign Language such as 'Enjera' 'Habesha' 'Resa' and so on, has been updated to Ethiopian Sign Language to be more culturally appropriate.

9. Methodology

9.1. *General steps for conducting SLR*

The following steps [Figure 4] refer to the approach or process used to conduct the paper. It outlines the systematic steps and strategies employed to achieve the objectives of the seminar. This methodology provides a framework for organizing and implementing the various components of the manual and non-manual recognition of Amharic sign language. Three steps were used to categorize the main tasks performed during the SLR: planning, conducting, and drafting the review report. Indeed, planning the SLR requires first considering the precise inclusion and exclusion standards of several earlier investigations.

9.2. *Inclusion and Exclusion Criteria*

We also used inclusion and exclusion criteria to determine the eligibility of participants for a seminar. These criteria help us to ensure that the participants selected for the Amharic sign language meet specific requirements and are representative of the target audience or objectives.

9.3. *Primary Data Selection Process*

Finding a primary study started by finding a database. From this, we use formulated search strings or keywords and apply exclusion criteria. Consequently, from our

Fig. 6. Step for Conducting SLR [18].

Table 1. Exclusion Criteria

Exclusion Criteria
1. Studies unrelated to Sign Language
2. Non-peer-reviewed publications or grey literature
3. Opinion pieces, reviews, or theoretical papers
4. Experimental studies, simulations, or case studies
5. Mid-night video acquisition and recognition
6. Research of sign Language related to Lexical and Syntactical analysis
7.Study related to Morphological analysis
8. Research conducted in Other than English Language

Table 2. Inclusion Criteria

Inclusion Criteria
1. Studies focusing on Amharic Sign Language detection and Recognition
2. Studies about deep learning techniques
3. Research conducted in AI, DL, AND/OR ML
4. Publications in peer-reviewed journals
5. Studies published in the last 5 years
6. Full-text articles available in English
7. Experimental studies, simulations, or case studies
8. Amharic Sign Language studies

study, we found 107 and when we applied exclusion criteria, only 67 papers were left. We have also used manual search to meet our criteria. After applying inclusion and exclusion again we finally got 41 papers that match our keywords. From 41 papers: IEEE Explore 17, Elsevier 5, Google Scholar 15, and Springer 4.

Table 3. Data Selection criteria

Primary Data Selection Process
Step1: Selection of research repository
IEEE Xplore (3828)
ScienceDirect (329)
Google Scholar (15900)
Springer (144)
ACM (4556)
Step 2: Applying exclusion criteria
Step 3: Initial selection of primary studies
Step 4: Applying inclusion criteria
Step 6: Applying inclusion/exclusion criteria again

10. Search Strategy and Source Selection

Developing an effective search strategy is crucial for conducting comprehensive and focused research. Hence, based on the above direction and our study objective, we selected the following database source. These are:

- IEEE Explore
- Elsevier,
- Google Scholar,
- Springer,
- ACM Digital Library, and so on

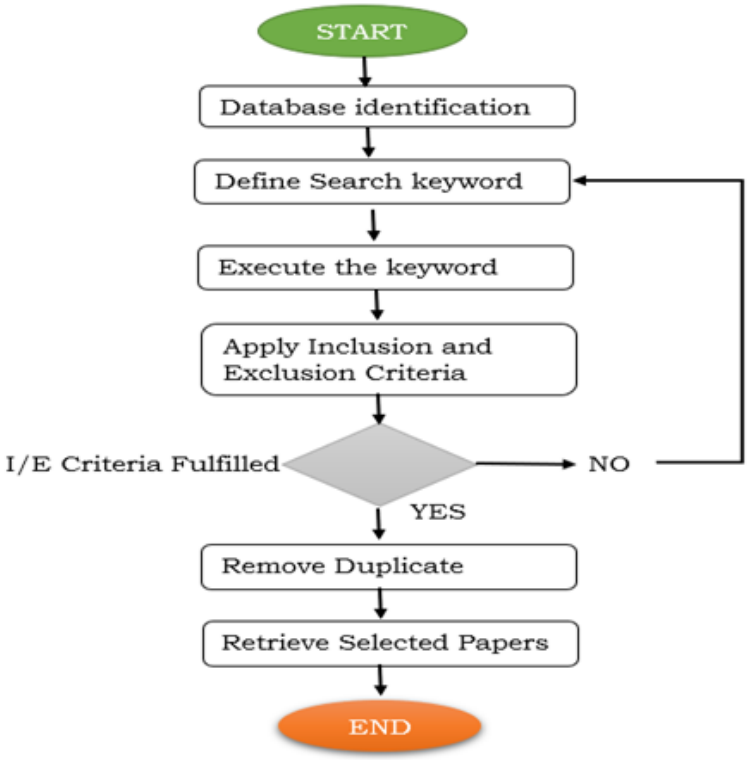

Fig. 7. Primary Data Selection Process

For the appropriate, reputable, and reputable selection of recent papers, one must use keywords. Since we are eager to systematic literature review on Amharic sign language, we defined ourown keywords that are listed below with operators "OR", "NOT" and "AND".

10.1. *Data Count Before and After Selection Process*

The following Table emphasizes the number of primary studies found before and after a selection process from reputable journals.

Quality assessment in a Systematic Literature Review (SLR) is a critical step that involves evaluating the methodological rigor and quality of the studies included in the review. This process is essential for ensuring that the findings and conclusions drawn from the review are based on sound, reliable, and valid evidence. The quality assessment Ensures Validity and Reliability, Reduces Bias, and Enhances Credibility. Hence, we have used the quality assessment question to justify review is fulfilled.

Table 4. Selection Keywords

(Deep Learning OR Deep Learning Approach OR Deep Neural Network OR Deep Learning Algorithm OR Deep Reenforcement learning) AND (Detection OR Recognition) AND (Artificial Intelligence) AND (Machine Learning OR Models) (Sign Language AND Recognition OR Detection OR Translation) (Amharic Sign Language AND Detection OR Recognition OR Detection) AND (Ethiopian Sign Language) AND (Amharic Sign Language) NOT (Speech)

Table 5. Data Count before and after selection

Source	Before Selection	After Selection
IEEE Explore	3,828	17
Elsevier	329	5
Google Scholar	15,900	15
Springer	144	6
ACM	4556	3
Total	24,757	46

Table 6. Quality assessment

1	What is the specific research objective or question addressed in the study?
2	Are the research methods and techniques employed clearly described?
3	How well is the concept of SLR explained?
4	What types of algorithms are considered in the study?

10.2. *Data extraction form*

We have also applied a data extraction form that is typically used when conducting research or gathering information from participants. It is a structured tool that helps us systematically collect and record data relevant to the seminar's objectives in the case of sign language recognition.

Table 7. Data extraction form

All types of focus (SLR)	Paper Content	Description
	Bibliography	Authors, Title, Year, Source
	Article types	Journal articles, Science Direct, IEEE Explore, Elsevier, ACM Digital Library, Scopus, Web of Science

11. Research Finding

The main aim of this systematic literature review is to answer the research questions asked above by using selected research papers. We hereby answered and elaborated the research questions with the help of the identified study using the criteria we set for this systematic literature review.

11.1. *Research Question 1 (RQ1)*

11.1.1. *How can hand gesture recognition be improved?*

In human-computer interaction, hand-gesture recognition is a significant challenge [18]. Among different forms of hand gesture recognition, one form is static hand gesture recognition. Feature extraction Module, Processing Module, and Classification Module are types of static hand gesture recognition developed by authors. The feature extraction module employs a top-down approach to human pose estimation to extract not just the key points but also bounding boxes for the body and hands. In the processing module, after being normalized and handled, its output will serve as the source of information for the categorization module in their suggested design, an architecture with two pipes. Based on the character of the datasets, the feature extraction module uses MMDetection for detecting hand gestures or full-body gestures. MMDetection is an open-source deep learning toolbox developed by the Multimedia Laboratory at The Chinese University of Hong Kong. It is designed for object detection research and implementation. MMDetection [34] provides a flexible and modular framework for training and evaluating state-of-the-art models for various computer vision tasks, with a primary focus on object detection. Some of its key features include module design, support of various data sets, and multiple model implementation. The MMdetection was used to detect the whole-body bounding boxes on datasets which includes the image of whole or upper half-human body (WH) datasets. On the data sets, only hand bounding boxes are detected not the whole human body since it will become more complex and extremely difficult

to train the model.

The paper used a pose estimation method for the processing of hand gestures. The pose estimation has the advantage of not wearing any sensor during data gathering. After data gathering the preprocessing of the data is followed by the usage of key points. The use of key points or poses has an advantage over complex backgrounds and different lighting conditions as well as the distance between the camera and hand gesture. Experiments were carried out on three datasets: HANDS, OUHANDS, and SHAPE. On three datasets, the suggested two-pipeline architecture with 2.5 million parameters achieved accuracy of 94%, 98%, and 94%. Furthermore, the lightweight version with 0.22 million parameters obtained 91%, 94%, and 96% accuracy. According to the paper [35], a vision-based gesture recognition system for Indian Sign Language (ISL) was proposed to recognize:

- Single-handed static and dynamic gestures,
- Double-handed static gestures, and
- Fingerspelling words.

Signs are extracted from a real-time video using skin color segmentation. An appropriate feature vector is extracted from the gesture sequence after the co-articulation elimination phase. The obtained features are then used for classification using a Support Vector Machine (SVM). The system successfully recognized:

- Finger spelling alphabets with 91% accuracy and
- Single-handed dynamic words with 89% accuracy

The study entitled "Hand Gesture Recognition Based on Single-Shot Multi-box Deep Learning" [19] has a scope of recognizing hand gestures to increase effective communication between humans and computers. The paper proposes a good approach to recognizing hand gestures in a complex scene or indifferent environment using the Single-Shot Multi-box Detector algorithm. This algorithm has 19 layers and the data prepared is the benchmark of the database which is prepared by considering different backgrounds or for a real-time hand gesture recognition system. Finally, the algorithm accuracy is found to be 99.2 Adversarial learning techniques can be employed to address challenges and improve the accuracy of sign language recognition systems, especially in scenarios where the input data may be subject to variations, noise, or adversarial attacks. Feature extractor for RGB videos The motion trajectory of sign language is a crucial factor. For RGB videos, considering that 3D-CNNs can extract the information of the trajectory and the residual networks are successful in different computer vision tasks, we employ the ResNet-18 version of the R (2+1) D network as the feature extractor [20]. The author represented the feature output by the first fully connected layer as m. The dimension of m is c, which equals the number of channels. Then they sent m into the sigmoid function and output s as the weights of channels. Finally, the element-wise product between s and f 2.

(a) Face keypoint

(b) Hand keypoints

Fig. 8. Key points[18].

11.1.2. *Feature extractor for RGB videos*

The motion trajectory of sign language is a crucial factor. For RGB videos, considering that 3D-CNNs can extract the information of the trajectory and the residual networks are successful in different computer vision tasks, we employ the ResNet-18 version of the R (2+1) D network as the feature extractor [20]. The author represented the feature output by the first fully connected layer as m. The dimension of m is c, which equals the number of channels. Then they sent m into the sigmoid function and output s as the weights of channels. Finally, performed element-wise product between s and f^2:

$$s = sigmoid(m), u = s \otimes f^2$$

where u is the new feacher map sent into the next convolutional block.

By training the auxiliary task the network can automatically focus on-hand locations.

Table 8. Module setting of the hand location tracking task

Position of the Module	Accuracy (%)
Conv block 1	64.9
Conv block 2	66.65
Conv block 3	66.16
Conv block 3	66.04

11.2. *Research Question 2 (RQ2)*

11.2.1. *Which features of ASL provide better and more accurate recognition of sign language in ASL?*

The paper published by Yigremachew Eshetu [3] discusses a real-time Ethiopian Sign Language translation to audio converter with a hybrid of vision-based and sensor-based area unit to capture hand configurations and knowledge of the corresponding meaning of gestures. The paper uses the pointed method for the following purposes discussed separately. The first one is a vision-based approach: they use a single camera to track the user's hands to recognize Ethiopian Sign Language. The system uses a machine learning neural network called Single Shot Multi-Box Detector (SSD) on TensorFlow. This network helped them to detect the hand gesture. The main part of the steps for this vision-based sensor is face detection by using haar-cascade [17] [3]to justify whether there are who are going to sign, hand detection, and overlap detection. This overlap detection indicates that whenever the system takes a video from real-time it identifies and analyzes the video if there is an overlap of hands and face in Ethiopian Sign Language meaning. For example, if someone puts his hand on his chin this means in Ethiopian Sign Language 'Mother' and if someone puts his/her hands on his/her forehead this means 'Father'. This indicates that according to the method of SSD, hands, and forehead are overlapped so that there is Amharic Sign Language meaning depending on the place of hands the user located. The average recognition rate obtained is 88.56% across all the gestures and for each of those gestures, the typical recognition rate is found to be 80%-90%.

11.3. *Research Question 3 (RQ3)*

11.3.1. *What are the most widely used deep learning algorithms for signer-independent sign language recognition?*

There is various study done on sign language, particularly with a deep learning algorithm. Below will discuss most deep learning algorithms used for Amharic sign language recognition. For the hypothesizing of object location, there is a need for state-of-the-art which depends on the region proposal such as Faster R-CNN and SSD [8] [3]. There are so many state-of-arts for object detection such as R-CNN, Fast R-CNN, and YOLO (You Look Only Once). Each of them differs from each other in either the speed or accuracy of detecting an object. For example, the algorithm R-CNN consumes an extreme amount of time training than the rest of the state. Faster R-CNN evolved to solve the speed and accuracy of the previous version Fast R-CNN [17] [8] [21]. For recognition of Amharic Sign Language to Amharic characters using Faster R-CNN, the following general steps are followed.

- Input frame/image of Amharic Sign Language extracted from a video of the signer and passed to ConvNet; then get a feature map of the image.

- RPN network is applied on the feature map found, RPN returns object proposal with their score of Sign Language image or not.
- RoI pooling layer is applied to the returned proposal to resize all the proposals with the same size.
- The last step states that the proposal is passed to a fully connected layer which has the SoftMax layer and the regression layer at its top, to classify and output the bounding box for Sign Language already annotated.

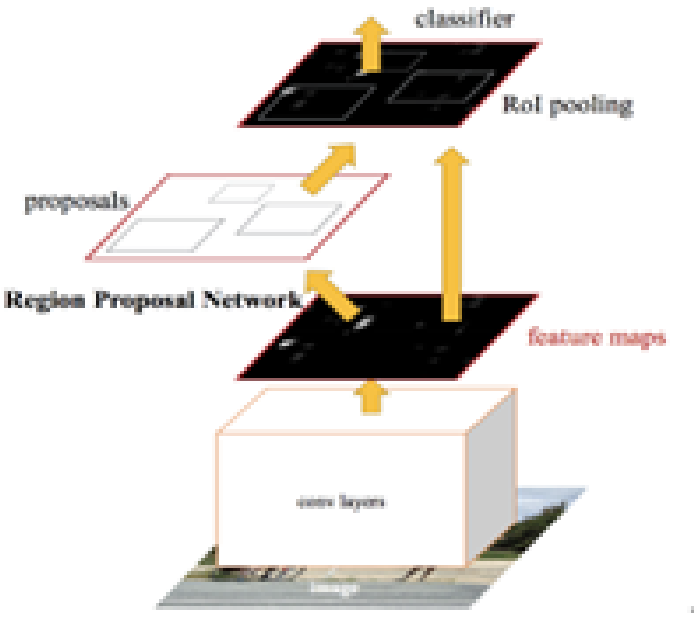

Fig. 9. Faster R-CNN reading frame [8]

12. Conclusion

Improving hand gesture recognition involves the finding of key points or poses i.e. finding the joint of the hand for effective training of the data. According to the review, there is no specific feature that is intended to provide better and more accurate recognition of ASL. This is because, with any feature of ASL with good data and the right technology we can get better recognition. The most widely used deep learning algorithm is CNN.

The systematic literature review on signer-independent, manual, and non-manual recognition of Amharic sign language highlights the existing research efforts and advancements in this domain. As we have seen in research questions 1, 2, and 3,

Table 9. Summary of Research Q3

Ref	Algorithm	Data Sets	Metrics
[4]	Single-Shot Detector (SDD)	Custom	88.46% accuracy
	Sensor-Based	Custom	88-90% accuracy
	Faster R-CNN	Prepare with d/t background	8.2 % and recognition time was 90.03 milliseconds
	faster R-CNN, SDD	Custom Prepared	98.25%, 96% accuracy

the review underscores the significance of developing accurate and robust sign language recognition systems that can effectively understand and interpret Amharic sign language gestures. The literature review revealed that various approaches have been explored for signer-independent recognition, aiming to overcome the challenges posed by variations in signing styles and individuals. These approaches include the utilization of depth sensors, computer vision techniques, and machine learning algorithms to capture and analyze manual and non-manual features of Amharic sign language.

The review also emphasized the importance of considering both manual and non-manual aspects of Amharic sign language, as non-manual features such as facial expressions, body posture, and head movements play a crucial role in conveying meaning in sign language communication. While significant progress has been made in manual and non-manual Amharic sign language recognition, the literature review identified some remaining challenges. These challenges include limited annotated datasets specific to Amharic sign language, the need for more comprehensive and standardized evaluation metrics, and the requirement for more extensive research on non-manual feature recognition.

The systematic literature review suggests that future research in this area should focus on addressing these challenges by collecting larger and more diverse datasets, developing robust feature extraction methods for both manual and non-manual cues, and exploring advanced machine learning techniques such as deep learning and adversarial learning for improved recognition performance. Overall, the systematic literature review highlights the growing interest in multimodal fusion to combine manual, and non-manual recognition of Amharic sign language and provides valuable insights for researchers, practitioners, and developers working to-

wards the advancement of technology-driven solutions for effective communication and inclusion of the Amharic signing community.

References

1. C. Zhu and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 3, pp. 569–573.
2. S. Chadha, A. Cieza, and E. Krug, "Global hearing health: future directions," *Bulletin of the World Health Organization*, vol. 96, no. 3, p. 146.
3. Y. Eshetu and E. Wolde, "A real-time ethiopian sign language to audio converter," *International Journal of Engineering Research & Technology (IJERT)*, vol. 8, no. 08.
4. S. Nagarajan, T. Subashini, and M. Balasubramanian, "Visual interpretation of asl finger spelling using hough transform and support vector machine," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 28–39, 2015.
5. Y. Wang, Y. Li, Y. Song, and X. Rong, "The influence of the activation function in a convolution neural network model of facial expression recognition," *Applied Sciences*, vol. 10, no. 5, p. 1897.
6. N. Yigzaw, M. Meshesha, and C. Diriba, "A generic approach towards amharic sign language recognition," *Advances in Human-Computer Interaction*, vol. 2022.
7. B. Belete, "College of natural sciences," Ph.D. dissertation, Addis Ababa University.
8. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28.
9. A. Wadhawan and P. Kumar, "Sign language recognition systems: A decade systematic literature review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 785–813.
10. S. Sarkar, B. Loeding, and A. S. Parashar, "Fusion of manual and non-manual information in american sign language recognition," in *Handbook of pattern recognition and computer vision*. World Scientific, 2010, pp. 477–495.
11. A. S. Al-Shamayleh, R. Ahmad, M. A. Abushariah, K. A. Alam, and N. Jomhari, "A systematic literature review on vision based gesture recognition techniques," *Multimedia Tools and Applications*, vol. 77, pp. 28 121–28 184.
12. H.-D. Yang and S.-W. Lee, "Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine," *Pattern Recognition Letters*, vol. 34, no. 16, pp. 2051–2056.
13. D. Z. Zeleke, "Amharic sentence to ethiopian sign language translator," Ph.D. dissertation, Addis Ababa University, 2014.
14. A. M. Gezmu, B. E. Seyoum, M. Gasser, and A. Nürnberger, "Contemporary amharic corpus: automatically morpho-syntactically tagged amharic corpus,"

arXiv preprint arXiv:2106.07241.

15. M. Tesfaye, "Machine translation approach to translate amharic text to ethiopian sign language."
16. S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1195–1215.
17. O. B. Hoque, M. I. Jubair, M. S. Islam, A.-F. Akash, and A. S. Paulson, "Real time bangladeshi sign language detection using faster r-cnn," in *2018 international conference on innovation in engineering and technology (ICIET)*. IEEE, pp. 1–6.
18. T. L. Dang, S. D. Tran, T. H. Nguyen, S. Kim, and N. Monet, "An improved hand gesture recognition system using keypoints and hand bounding boxes," *Array*, vol. 16, p. 100251.
19. P. Liu, X. Li, H. Cui, S. Li, and Y. Yuan, "Hand gesture recognition based on single-shot multibox detector deep learning," *Mobile Information Systems*, vol. 2019, pp. 1–7.
20. D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6450–6459.
21. M. Rivera-Acosta, J. M. Ruiz-Varela, S. Ortega-Cisneros, J. Rivera, R. Parra-Michel, and P. Mejia-Alvarez, "Spelling correction real-time american sign language alphabet translation system based on yolo network and lstm," *Electronics*, vol. 10, no. 9, p. 1035.