Grenze International Journal of Engineering and Technology, Jan Issue

Review Paper on Enhancing Communication: Machine Learning for Live Sign-to-Text Translation

Aditya Rangari[1], Devendra Bhide[2], Vaibhav More[3], Kunal Wahurwagh[4], Dhiraj Shirbhate[5] and Chetan Andhare[6]
[1-4]Student, Department of Computer Engineering, Government College of Engineering, Yavatmal, Maharashtra, India
[5]Assistant Professor, Department of Computer Engineering, Government College of Engineering, Yavatmal, Maharashtra, India
[6]Head of Department, Department of Computer Engineering, Government College of Engineering, Yavatmal, Maharashtra, India

***Abstract*— For those who are deaf or hard of hearing, sign language is essential as their main form of communication. using ease, sign language gestures may be translated into written or spoken words in real time using the Sign Language Translator, and vice versa. This system interprets and communicates sign language gestures by utilising computer vision and natural language processing (NLP). Given that sign language uses a wide range of hand movements to communicate meaning, it might be difficult to identify certain motions by looking for patterns. Individuals communicate and engage using a variety of gestures. In this study, a human-computer interface that can recognise motions in sign language and properly translate them into text is shown. The suggested method improves interpersonal communication by using convolutional neural networks and long short-term memory networks for gesture interpretation and detection.**

***Index Terms*— Language translator, Natural language processing (NLP), Human-computer interface, gestures, Convolutional neural network, Long-Short-term memory network, detection, etc.**

I. INTRODUCTION

Sign language, as a visual and gestural mode of communication, plays a pivotal role in the lives of millions of individuals who are deaf or hard of hearing, serving as their primary means of expressing thoughts, emotions, and ideas. Yet, despite its profound significance, the gap in communication between sign language users and the wider society remains a formidable barrier to inclusive social interaction, education, and employment opportunities. The evolution of technology, particularly the advent of machine learning and artificial intelligence, has presented an unprecedented opportunity to bridge this communication divide, offering a path to more seamless and effective communication between the deaf and hard of hearing communities and the hearing world. This review paper, titled " Enhancing communication: machine learning for live sign-to-text translation " embarks on a comprehensive journey through the landscape of this transformative field, which has seen remarkable advancements in recent years. Historically, sign language interpretation has largely relied on the expertise of human interpreters, a valuable but limited resource. The availability and affordability of skilled interpreters have often fallen short of the demand, hindering effective communication and access to essential services for sign language users.

Grenze ID: 01.GIJET.10.1.126

Furthermore, traditional approaches to sign language interpretation have faced challenges in capturing the nuances and intricacies of sign language, resulting in potential misinterpretations and communication gaps. Machine learning, with its ability to process and understand complex visual and linguistic data, has emerged as a promising solution to these long-standing challenges. By leveraging computer vision techniques for sign language recognition and natural language processing for text generation, machine learning-assisted live sign-to-text translation systems aim to provide real-time, accurate, and contextually rich interpretations of sign language. These systems hold the potential to not only enhance communication but also facilitate the integration of individuals who are deaf or hard of hearing into various aspects of society, including education, employment, and social engagement.
This review paper seeks to offer a comprehensive overview of the state-of-the-art techniques, methodologies, and innovations in the field of machine learning-assisted live sign-to-text translation. It begins by discussing the challenges faced by sign language users and the critical need for effective translation solutions. We will then trace the historical evolution of sign language interpretation methods, highlighting the limitations of traditional approaches that have motivated the exploration of machine learning solutions. The subsequent sections will delve into the core components of machine learning systems in this context, exploring the various techniques and models developed for sign language recognition and natural language text generation. A critical analysis of these models, their associated datasets, and performance metrics will be presented to provide a comprehensive understanding of their capabilities and limitations. Moreover, we will investigate the technologies and tools that enable live sign-to-text translation, including wearable devices, computer applications, and mobile applications, each with its own unique features and capabilities. Techniques such as data augmentation, transfer learning, and real-time processing will also be explored for their roles in enhancing the accuracy and speed of translation systems. In addition to the technical aspects, this review will also consider the user experience and societal acceptance of machine learning-assisted sign-to-text translation systems. Human factors and accessibility considerations will be examined, shedding light on the practical implications of these technologies in real-world scenarios. Ethical considerations, potential biases, and challenges associated with machine learning-assisted sign language translation will be discussed, emphasizing the importance of designing these systems with fairness and inclusivity in mind. Data is collected using a real-time system identification method created by Sakshi Goyal and Ishita Sharma (2015). A feature extractor then divides the data into different frames and characteristics, such as the Gaussian difference in Centroids [1]. Between 250 million to 300 million people around the world have speech and hearing difficulties. Sign languages exhibit distinct characteristics and are not universally comprehensible, despite apparent commonalities among them. Sign languages are prevalent in various nations, with examples including Sri Lankan Sign Language, American Sign Language, British Sign Language, Indian Sign Language, and French Sign Language. The current study investigates a live translation system designed for Indian Sign Language, enabling the conversion of English text into Indian Sign Language and vice versa, also providing options for text or audio formats. The primary objective of this research is to promote improved communication between the general public and individuals with hearing or speech impairments in India, empowering them to independently learn sign language [2].

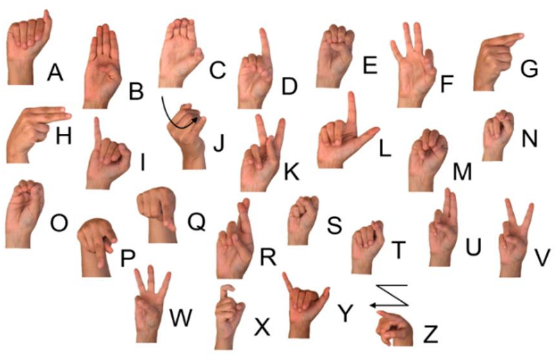

Figure 1: Sign Language in Indian

The majority of research on gesture recognition has concentrated on modelling the multichannel information about hand gestures—hand movement and shape—that is extracted from the visual signal in order to recognise signals. As a result, several strategies, including hidden Markov models (HMM), have been created over time for the machine learning methods required to understand sign language. [4-6], boosting [9], sequential pattern trees [10], boosting [7], parallel HMM [7], relevance vector machines [8], and deep learning techniques [11, 12].

The field of sign language recognition technology is still relatively new, even with recent breakthroughs. Resource limitation is one of the main causes, along with the challenge of gathering and processing data from several sources. Sign languages are less widely used than spoken languages because of this. Furthermore, the lexicon and syntax of sign languages are different from those of their corresponding spoken languages [1]. For example, British sign language does not translate to British English. Furthermore, there's a chance that the sign languages are different even though the spoken languages might be the same. For example, the sign languages used in America and Britain are different. German Sign Language (DGS) and Swiss German Sign Language (DSGS) can be distinguished similarly. The resource scarcity conundrum can be resolved by developing systems that can make use of multiple sign language resources by overcoming the limitations imposed by the differences between the sign languages. To be more precise, that approach just incorporates hand form modelling, and there isn't much research on it in the literature. Combining resources from many sign languages can be used to train a global hand shape classifier. Given the HamNoSys annotation [13] of generated signals, hand shape information-based sign language recognition systems can be developed [14].

II. Literature Review

Throughout my research, I came across a number of books that focused on the Translation System for Dumb and Deaf People and all of its components and methods. Data is collected using a real-time system identification system created by Sakshi Goyal and Ishita Sharma (2015). A feature extractor then divides the data into different frames and features, such as the Gaussian difference in Centroids. [1] Iker Vazquez Lopez created the programme Language Transcription in 2017, which interprets hand movements in images using algorithms for image identification and analysis. To recognise motions, the processes of hand placement, hand segmentation, and classification are applied. [2] Mr. Vedant D. Shinde and Professor Radha S. Shirbhate (2020) created a mechanism to construct a Sign Language Classification employing several computer vision algorithms, such as SVM and KNN, to generate an autonomous sign language recognition system that is used in real-world situations with a range of instruments. [3] Mohammad Elham Walizad and Mehreen Hurroo (2015) developed a Signs Language Recognition system using Convolutional Networks and machine vision. They used segmentation techniques, splitting to evaluate the entire region of skin tone, and there was no discernible difference between photographs of different motions because OpenCV lowers all of the pictures to the same size. [4].

Studies on sign languages are carried out all around the world. American Sign Language, Chinese Sign Language, and Sinhala Sign Language are a few of them [1]. These sign language detections employ two methods. While the other is image- or vision-based, the first is device-based [2][3]. When identifying signs, preprocessing, feature extraction, and classification are usually carried out in a single simple step. Pre-processing is done to take an undesirable component out of the picture. The Region of Interest (ROI) will have its image features extracted and class

A. Self-sufficient Sign Language Recognition in India: Handling Co-articulation in Real-time Video Broadcasts [16]

Without the need for extra sensors to detect hand areas, the authors' technique can recognise ISL motions from films taken using a mobile device's camera. This system recognises single-handed static and dynamic motions as well as double-handed static gestures. The ROI method is used by the system to extract features. The centroid of the binary image is located following pre-processing. Both static and dynamic motions are recognised by the gesture separation technique. This system is very user-friendly due to its low cost and simplicity of installation, both of which can be accomplished with a mobile camera. Its inability to function properly in dimly lit environments or against busy backdrops is a drawback.

B. A Deep Learning Framework Customised for the Recognition of Indian Sign Language [17]

This system's testing and validation accuracy was 98.64%, while its training accuracy across all ISL static alphabets was 99.93%. Facial expression analysis and situation analysis are the two components of this study that are left out.

C. Indian Two-Handed Sign Language Instant Recognition Using Binarized Neural Network Approach [18]

Considering the challenges of gesture identification focused on specific embedded systems, the authors have presented a novel design of a binarized neural network using bitwise operations and binary variables for activations and weights. The advantage of this design is that it achieves 98.8% total accuracy, which is higher than prior systems. The disadvantage is that some signals of M, N, and E are misclassified by this system due to their similar shapes. Furthermore, the limited number of gesture classes imposes constraints on the proposed BNN design.

D. Three-Dimensional CNN Hand Gestures for Human Sign Language Recognition [19]

This article investigates the use of 3DCNN for hand gesture recognition. The authors compared their suggested strategies with six additional state-of-the-art methods from the literature. Four of these they outperformed, and their performance was comparable to the other two ways. But it doesn't work for a live video stream.

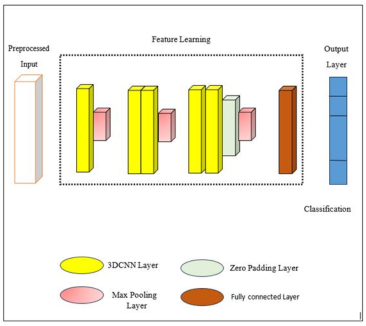

Figure 2.1: A single structure built on 3DCNN

E. Detection of American Sign Language using RF Sensing Method [20]

This work provides a comprehensive, interdisciplinary perspective on RF sensor-based ASL detection. The findings show that RF sensing can effectively function in the dark and preserve user privacy while offering contactless ASL identification for ASL-sensitive smart environments. In order to offer significant interpretations for technology and algorithm correlation, a large multimodal database of connected native signatures would be required.

F. Applying Video-sourced Deep Cascading Models Separate Hand Sign Recognition [21]

Using CNN, LSTM, and SSD, the authors' deep learning model can systematically identify hand signals from RGB videos. This strategy enhanced the hand sign recognition accuracy and intricacy. It allowed for quick processing in an uncontrollable environment, including quick hand movements. Accurate detection may be improved by gathering more data.

G. Relativistic depth Method for Decoding Indian Sign Language Using Microsoft Kinect [22]

This method was effective in identifying overlapping signs, double hand signs, and signs unique to ISL. The project's advantage is that the average recognition accuracy increased by 71.85% when this method was used. The system reached 100% accuracy for a small number of signals; however, it often generates erroneous translations due to its disregard for the surrounding context of the motions.

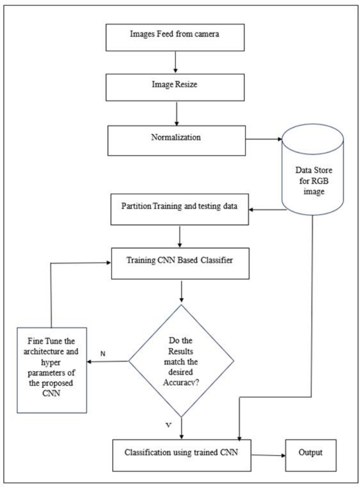

Figure 2.2: Proposed System

H. Continuous Sign Language Recognition Making Use of Leap Motion and a Modified LSTM Method [23]

In this study, they have created a cutting-edge architecture for continuous-SLR using the Leap motion sensor. An adapted LSTM architecture has also been proposed for sign language recognition. It was discovered that the average accuracy for both solitary sign words and signed phrases was 89.5% and 72.3%, respectively. More training data will lead to improvements in both recognition and model learning performance.

I. Sign Language Interpreter Using Fixed Gestures and Deep Learning [24]

With respect to differences in the number of layers and filters, 99.17% training accuracy and 98.80% validation accuracy were the greatest values that the authors achieved. The identifying method has to be improved by gathering more datasets.

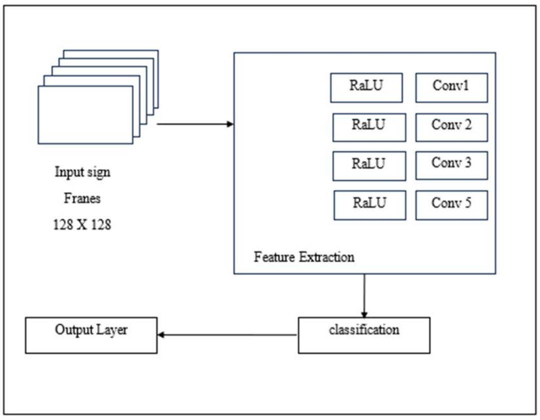

Figure 2.3: Proposed Deep CNN Structure

J. Deep Neural Network-based Recognition of Sign Language [25]

For the purpose of classifying movements in selfie sign language, the authors proposed a CNN architecture. Although there are less documented training and validation errors with the proposed CNN design, the database is not made publicly accessible.

III. METHODOLOGY

The product is comprised of four main components: "Detecting Hand Gestures," "Visual Recognition System," and "Converter for Text to Sign"—all of which work together to give an efficient translation tool for all Indians.

A. Analysing and Identifying Hand Movements

This component will be developed using a Faster RCNN (Region Based Convolutional Neural Networks) setup on top of the TensorFlow model trainer. For the model training, 247 images, or three pictures for every letter in the English-based Indian Sign Language alphabet, were employed. First, a label map is created for each of the classes concerned. In this case, there is only one class—the hand. 800 by 600-pixel resolution will be used for the photos in order to optimize storage efficiency and training pace. Every photo will be captured using low-quality laptop webcams. The aim is to acquire low-resolution images for the basis of the training process. Therefore, when the model is run, it will be identified even if it is fed a low-pixel image. Low latency network connections, where less uplink and downlink network traffic is needed, would also benefit from this. Indians who reside in rural regions with sluggish internet access can profit from this.

The identified pictures must be divided into train and test datasets in order for the model to be trained effectively. After that, an indicating tool in Pascal VOC format is used to label the photos, creating an extensible markup. Every picture has a language record in XML. Those individual photos are converted into CSV files for the train and test image datasets. The model is first trained using the TensorFlow model trainer and Faster RCNN Configuration using CSV data. The RCNN pet authentication mechanism that operates more quickly: this method is the most similar the human hand and tone provided the basis for this arrangement in terms of colour. Soon after a successful training session, the camera is used with the pre-trained model. to recognize pictures. A picture will be sent over an API to the Illustration Classifier element for further classifications if it has been identified with more than 95% accuracy.

B. Strategies for Image Categorization

An "Image Classifier" used in a Sign to Text converter is a critical component of the technology, tasked with the challenging job of recognizing and interpreting sign language gestures captured through images or video. This image classification process is pivotal in converting sign language into written or spoken text. To operate effectively, the classifier undergoes a training phase using a sizable dataset of sign language gestures, each labelled with the corresponding words or phrases. Machine learning algorithms, often convolutional neural networks (CNNs), are then employed to train the classifier. During this training process, the classifier learns to discern intricate patterns, shapes, and features within the sign language gestures.

These features, which may include handshapes, movements, and facial expressions, are crucial for distinguishing between different signs and conveying their intended meanings. Once trained, the image classifier can accurately recognize and classify sign language gestures, associating each gesture with the corresponding text, thus converting the visual information into a more accessible and easily understandable written or spoken format. This process is fundamental in making sign language communication accessible to a broader audience, promoting inclusivity and facilitating real-time conversations between Deaf individuals and those who may not be proficient in sign language. Additionally, the classifier's adaptability allows it to recognize various sign languages, making the technology applicable across different regions and cultures, ultimately enhancing accessibility in education, healthcare, customer service, and everyday social interactions.

C. Text-to-Sign Transformation Methods

A "Converter for Text to Sign" is a sophisticated technological tool designed to facilitate communication between individuals who use sign language as their primary means of expression and those who may not be proficient in sign language. This tool leverages natural language processing (NLP) and machine learning algorithms to analyse written or spoken text and generate corresponding sign language representations. Its primary objective is to make information and communication more accessible and inclusive for Deaf and hard of hearing individuals.

The core components of a Text to Sign converter typically include NLP algorithms, a sign language database, and potentially gesture recognition through computer vision. NLP techniques are employed to understand the linguistic content of the input text, encompassing sentence structure, grammar, and vocabulary. The sign language database stores the vocabulary and grammar rules necessary to generate the appropriate signs for each word or phrase. In some advanced systems, computer vision is utilized to recognize and generate sign language gestures, enhancing the accuracy and expressiveness of the signs.

These converters have a pivotal role in promoting accessibility and inclusivity, as they enable Deaf individuals to engage in conversations, access educational materials, news, literature, and online content. However, challenges include accurately representing the complex and region-specific nature of sign languages, achieving real-time translation with minimal latency, and ensuring cultural sensitivity in sign generation. The future of Text to Sign conversion technology is promising, with the ongoing advancements in machine learning and NLP leading to more accurate, responsive, and user-friendly tools that can significantly improve communication and accessibility for the Deaf and hard of hearing community. The integration of standardized sign language datasets and gesture recognition through technologies like webcams is expected to further enhance the quality and applicability of these converters, making them increasingly accessible across various platforms and applications.

IV. Description And Future Scope

The review article "Machine Learning-Assisted Live Sign-to-Text Translation" offers a thorough examination of the status of research and development concerning machine learning approaches for live sign language translation. The Deaf and hard of hearing community uses sign language as their primary form of communication, therefore applying machine learning to this industry has the potential to completely transform accessibility and inclusion in a variety of industries. The article provides a thorough overview of the present state of the subject by thoroughly reviewing the technology breakthroughs, obstacles, ethical issues, and user-centric design concepts used in these systems. The aforementioned underlines the critical role that this technology plays in closing gaps in communication and emphasizes how important it is to improving accessibility for Deaf and hard of hearing individuals.

Live sign-to-text translation with machine learning assistance has a bright future ahead of it. Higher precision and real-time processing will probably be the main goals of this field's research, which will make these systems even more useful for daily usage. Furthermore, the use of multimodal translation—which combines text, speech, and sign language—can offer a more complete communication solution that will be advantageous to a larger audience.

Additionally, the development of standardized sign language datasets and the establishment of privacy and ethical guidelines are expected to be key areas of focus. As this technology evolves, it will also require a user-centric approach, involving the Deaf and hard of hearing community in the design and testing of these tools to ensure that they are user-friendly and effective. In the coming years, the international adaptation of these systems, accommodating various sign languages and cultural contexts, will further enhance their global impact. Overall, the future scope of this field is filled with opportunities for innovation, collaboration, and the potential for a more inclusive and interconnected world.

V. Conclusion

Significant advancements in machine learning-based sign language recognition and translation have been made in response to the demand for accessible communication solutions. The literature review highlights how important developments in the area have been made and highlights the potential of Sign Language Converters to improve inclusiveness for the Deaf and Dump communities by removing communication obstacles. Further research is needed to address existing problems and improve these systems' usability and effectiveness in real-world scenarios. Thanks to a technique developed as part of this study, there shouldn't be any obstacles in the way of our communicating with stupid or deaf people. Finding the proper categorization is the goal of the convolution neural network.

References

[1] Sakshi Goyal, Ishita Sharma, Shanu Sharma 1“Sign Language Recognition System for Deaf and Dumb People”, International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 4, April – 2013.

[2] Iker Vazquez Lopez-Hand gesture recognition for sign language transcription, Boise State University, Research and Economic Development, 2017.

[3] Prof. Radha S. Shirbhate1, Mr. Vedant D. Shinde2, Ms. Sanam A. Metkari3, Ms. Pooja U Borkar4, Ms. Mayuri A. Khandge5 “Sign language Recognition Using Machine Learning Algorithm”, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 03 | Mar 2020.

[4] Mehreen Hurroo, Mohammad ElhamWalizad Sign Language Recognition System using Convolutional Neural Network and Computer Vision”, International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 12, December-2020

[5] T. Starner, J. Weaver, and A. Pentland, “Real-time American sign language recognition using desk and wearable computer-based video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371–1375, 1998.

[6] Oscar Koller, Necati Camgoz, Hermann Ney, and Richard Bowden, “Weakly supervised learning with multi-stream CNN-HMM to discover sequential parallelism in sign language videos,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 04 2019.

[7] S. Tornay, M. Razavi, N. C. Camgoz, R. Bowden, and M. Magimai. -Doss, “HMM-based approaches to model multichannel information in sign language inspired from articulatory features-based speech processing,” in Proc. of ICASSP, 2019.

[8] C. Vogler and D. Metaxas, “Parallel hidden Markov models for American sign language recognition,” in Proc. of the Seventh IEEE International Conference on Computer Vision (ICCV), Sep. 1999, vol. 1, pp. 116–122 vol.1.

[9] G. Awad, J. Han, and A. Sutherland, “Novel boosting framework for subunit-based sign language recognition,” in Proc. Of the 16th IEEE International Conference on the Image Processing (ICIP).

[10] E.-J. Ong, H. Cooper, N. Pugeault, and R. Bowden, “Sign language recognition using sequential pattern trees,” in Proc. of the IEEE CVPR, 2012.

[11] O. Koller, O. Zargaran, H. Ney, and R. Bowden, “Deep sign: hybrid CNN-HMM for continuous sign language recognition,” in Proc. of the British Machine Vision Conference (BMVC), 2016.

[12] N. C. Camg¨oz, S. Hadfield, O. Koller, and R. Bowden, “Subunets: End-to-end hand shape and continuous sign language recognition,” in Proc. of the IEEE International Conference on Computer Vision (ICCV), 2017.

[13] T. Hanke, “HamNoSys - representing sign language data in language resources and language processing contexts,” Workshop proceedings: Representation and processing of sign languages, pp. 1–6., 2004.

[14] O. Koller, H. Ney, and R. Bowden, “Deep hand: How to train a CNN on 1 million hand images when your data is continuous and weakly labelled,” in Proc. of the IEEE CVPR, June 2016.

[15] P. K. Athira, C. J. Sruthi, A. Lijiya, “A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: Scenario”,2019.

[16] M. Jaiswal, V. Sharma, A. Sharma, S. Saini, R. Tomar, “An Efficient Binarized Neural Network for Recognizing Two Hands Indian Sign Language Gestures in Real-time Environment”,2020

[17] C. J. Sruthi, A. Lijiya, “Signet: A Deep Learning based Indian Sign Language RecognitionSystem”,2019.

[18] T. Raghuveera, R. Deepthi, R. Mangalashri, R. Akshaya, “A depth-based Indian Sign Language recognition using Microsoft Kinect”,2020.
[19] A. Wadhawan, P. Kumar, “Deep learning-based sign language recognition system for static signs”,2020.
[20] R. Rastgoo, K. Kiani, S. Escalera, “Video-based isolated hand sign language recognition using a deep cascadedmodel”,2020.
[21] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford, M. M. Rahman, E. Kurtoglu, R. Aksu,
[22] T. Macks, R. Mdrafi, “American Sign Language Recognition Using RF Sensing”, 2020.
[23] M. Al-Hammadi, G. Muhammad, w. Abdul, M. Alsulaiman, M. A. Bencherif, M. A. Mekhtiche, “Hand Gesture Recognition for Sign Language Using 3DCNN”, 2020.
[24] G. A. Rao, K. Syamala, P. V. V. Kishore, A. S. C. S. Sastry, “Deep Convolutional Neural Networks for Sign Language Recognition”, 2018.
[25] A. Mittal, P. Kumar, P. R. Roy, R. Balasubramanian, B. B. Chaudhuri, “A Modified-LSTM Model for Continuous Sign Language Recognition using Leapmotion”,2019.
[26] NLTK.org, “Natural Language Toolkit.” https://www.nltk.org/ (accessed Jul. 01, 2020).
[27] tutorialspoint, “Natural Language Processing – Semantic Analysis.” https://www.tutorialspoint.com/natural_language_processing/natural_language_processing_semantic_analysis.htm (accessed Feb. 02, 2020).