Real-time Bangla Sign Language Translator

Rotan Hawlader Pranto
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
rotan.pranto@northsouth.edu

Dr. Shahnewaz Siddique
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
shahnewaz.siddique@northsouth.edu

***Abstract*—Abstract— The human body communicates through various meaningful gestures, with sign language using hands being a prominent example. Bangla Sign Language Translation (BSLT) aims to bridge communication gaps for the deaf and mute community. Our approach involves using Mediapipe Holistic to gather key points, LSTM architecture for data training, and Computer Vision for real-time sign language interpretation with an accuracy of 94%.**

***Keywords*—Recurrent Neural Network, LSTM, Computer Vision, Bangla font**

I. INTRODUCTION

Communication is essential for expressing feelings, yet it poses significant challenges for the deaf community. Globally, there are approximately 466 million deaf individuals, including 36 million children. In Bangladesh alone, around 13.7 million people are deaf. Disabilities affect approximately 16% of the world's population, underscoring the importance of inclusive communication solutions.

Sign language is the most efficient way to make communication between deaf and dumb people. However, this is only possible if both of them have an acute knowledge of sign language. Each word and alphabet has specific signs. So, prior knowledge is needed to facilitate convenient communication. Fig 1 shows different signs which represent the alphabet of the Bangla Sign language.

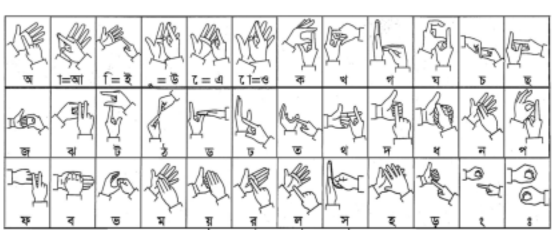

Fig. 1. Bangla alphabet sign [7]

Deep learning encompasses diverse methodologies, among which RNNs (Recurrent Neural Networks) stand out for their effectiveness in processing sequential data. LSTM (Long Short-Term Memory) networks, a specialized type of RNN, are particularly renowned for their ability to manage and predict data sequences over time. In the context of BSLT (Bangla Sign Language Translation), LSTM networks are used for training due to their proficiency in handling temporal patterns. The system integrates computer vision techniques for real-time sign language detection, utilizing the Mediapipe library to collect and track key points and landmarks on the user's hands and face. The PIL (Python Imaging Library) renders fonts, ensuring clear and accurate visual representation of translated text. By combining these advanced frameworks and libraries, BSLT efficiently translates Bangla sign language into written text, facilitating user communication and enhancing accessibility.

II. RELATED WORKS

Previously a lot of work has been done to recognize and translate sign language using various methods and technologies using different types of sensors. Among them, some relevant works are explained below.

A. Neural Sign Language Translation [1]

This paper offers a concise overview of sign language translation methodologies. The researchers utilized the PHOENIX-weather 2014T dataset to train their model. They achieved an upper bound translation performance of 19.26 BLEU-4. Their end-to-end frame-level and gloss-level tokenization networks attained respective scores of 9.85 and 18.13.

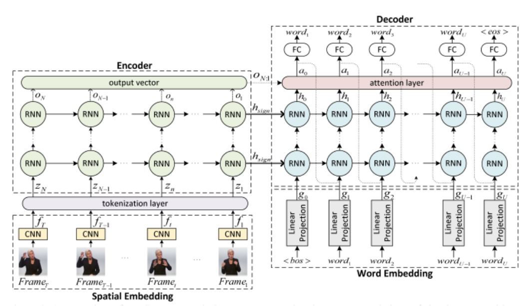

Fig. 2. An overview of our SLT approach that generates spoken language translations of sign language videos. [1]

B. Real-time Sign Language Recognition using Computer Vision [2]

Their paper addresses the societal gap between differently-abled individuals, such as those who are deaf and mute, and others. Image processing techniques are employed to preprocess images and extract hands from backgrounds effectively. The researchers utilized CNN to evaluate their custom dataset and real-time hand gestures, achieving an accuracy of 83%.

C. *Research of a Sign Language Translation System Based on Deep Learning [3]*

This paper explores hand localization and sign language recognition using neural networks. The approach integrates faster R-CNN for sign detection, 3D CNN for feature extraction, and LSTM networks for sequence encoding and decoding. The study achieved a notable 99% accuracy in recognizing common vocabulary datasets for sign language.

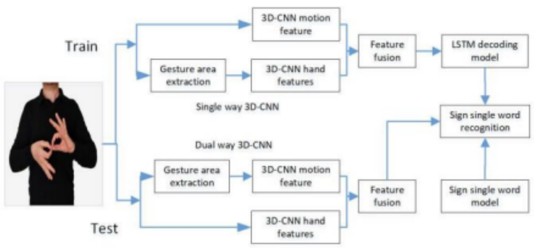

Fig. 3. Double-way 3D CNN and LSTM Encoding and Decoding Network Structure[3]

D. *Continuous Sign Language Recognition with Correlation Network [4]*

This paper (CorrNet) improves CSLR by capturing body trajectories across frames using a correlation module. It achieves state-of-the-art accuracy on datasets including PHOENIX14, with a Word Error Rate of 18.8% on the development set and 19.4% on the test set. CorrNet's effectiveness is demonstrated through comprehensive comparisons and visualizations, highlighting its ability to emphasize human body trajectories across adjacent frames.

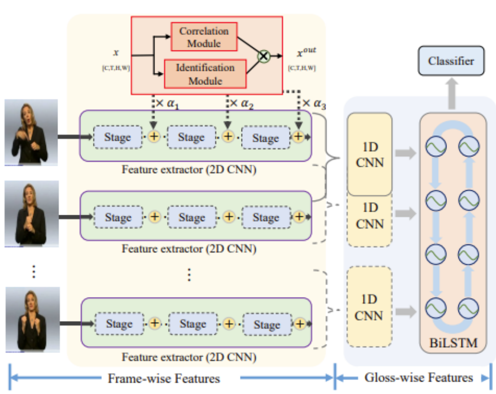

Fig. 4. An overview of CorrNet [4]

E. *Sign Language Translator and Gesture Recognition [5]*

The paper explores a sign language translator and gesture recognition system. The system leverages computer vision and machine learning techniques to accurately translate sign language and gestures. They have built a smart glove to detect hand movements. The proposed model demonstrates an accuracy score of 96%, showcasing its effectiveness in facilitating communication for individuals who use sign language.

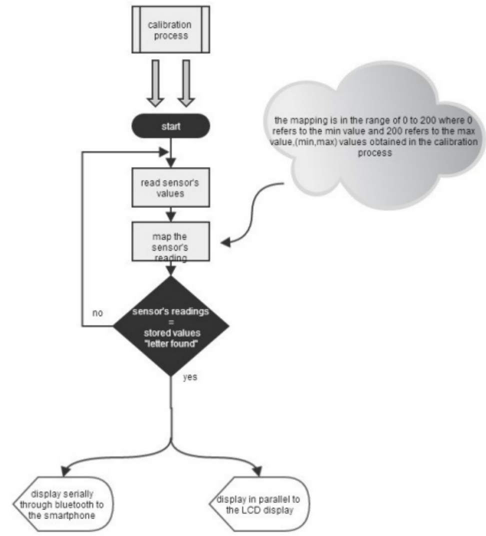

Fig. 5. The testing procedure flowchart [5]

III. PROPOSED SOLUTION

A. *Recurrent neural network*

Recurrent Neural Networks (RNNs), particularly those featuring Long Short-Term Memory (LSTM) architecture, are renowned for their ability to handle and generate sequential data. This makes them highly useful in applications such as sign language translation, where the sequential nature of gestures and signs are crucial. BSLT leverages LSTM networks for both encoding and decoding sequences of sign language gestures, allowing for the formulation of coherent and contextually accurate sentences. The system's integration of real-time sign detection ensures that the most recent and accurate signs are used to generate translations. By maintaining context over time and effectively processing sequential data, BSLT significantly enhances meaningful communication for individuals who use sign language, bridging gaps and facilitating better interactions. The advanced capabilities of LSTM networks in managing temporal dependencies is instrumental in achieving high accuracy and reliability in Bangla sign language translation.

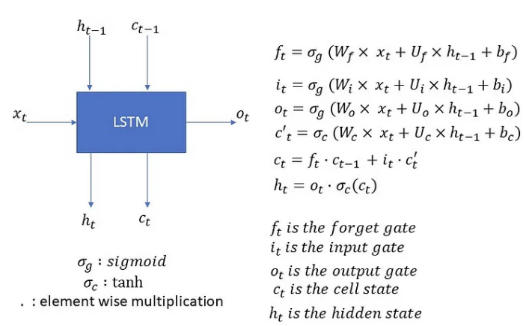

Fig. 6. Long short-term memory [11]

B. Methodology

We utilized Mediapipe for extracting key points and landmarks detection from images. The collected data was processed by flattening the key points into NumPy arrays and appropriately labeled. Our model employed an LSTM architecture to achieve high accuracy in generating coherent sentences. Additionally, we utilized the PIL library to render Bangla fonts, ensuring linguistic accuracy. Real-time execution was facilitated through computer vision.

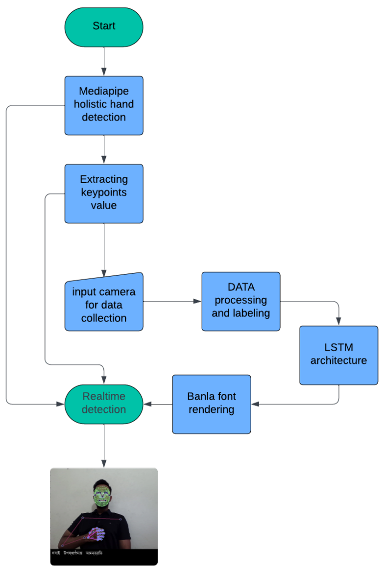

Fig. 7. Methodology

C. Data Collection

We systematically collected sequential data for each word, capturing 30 frames per word using MediaPipe Holistic for comprehensive landmark detection, including hands, face, and pose. Key points were extracted and flattened into NumPy arrays, ensuring a structured representation of sign language gestures for effective processing and analysis og The Bangla sign language translator.

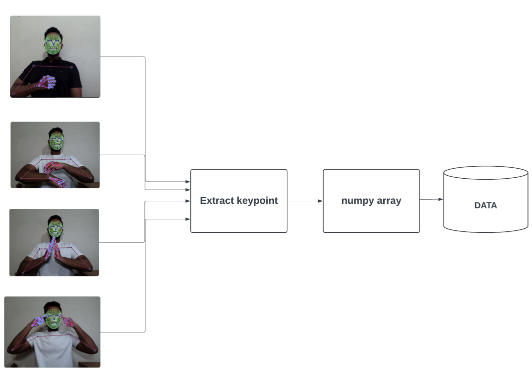

Fig. 8. Data collection

D. Results

Collecting data for Bangla Sign Language presented significant challenges, particularly due to the need for accurate and diverse representation of signs. Despite these difficulties, we successfully gathered and meticulously labeled the data, capturing 30 frames per word to ensure comprehensive coverage. We employed a Long Short-Term Memory (LSTM) architecture for training, achieving a commendable accuracy score of 94% and an F1 score of 93%. The model's performance was further validated through various tests, demonstrating its effectiveness in real-world scenarios. Our results include a detailed presentation of the model's predictions, as well as a screenshot showcasing real-time detection capabilities. These demonstrations highlight the practical utility of the model, offering a robust solution for real-time sign language recognition. The successful implementation of this model illustrates its potential for enhancing communication and accessibility for Bangla sign language users.

TABLE I
IN THIS TABLE, WE HAVE SHOWN THE RESULTS

Architecture	Accuracy	F1 Score
LSTM	94%	93%

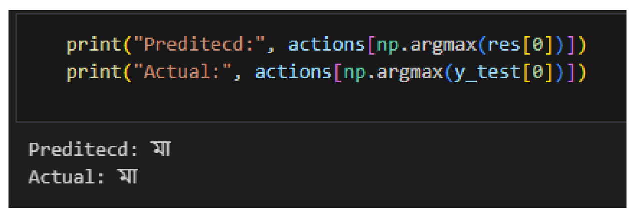

Fig. 9. Predicted result

Fig. 10. Realtime Detection

E. Experimental setup

We have utilized several libraries and tools for our project, including OS, OpenCV (CV2), NumPy, MediaPipe, Matplotlib, scikit-learn, TensorFlow, LSTM, TensorBoard, and metrics. Our data collection involved real-time sign detection using a camera. The hardware configuration comprises an AMD Ryzen 7 5700X CPU, a GeForce RTX 3060 OC 12GB GPU, 16 GB of 3200 MHz RAM, and an MSI B550 GEN3 motherboard.

IV. ANALYSIS

The primary challenge we encountered was the absence of a comprehensive dataset for Bangla words, which initially made data collection difficult. Specifically, distinguishing between the signs for "deer" and "educated" presented significant detection challenges. We resolved these issues before beginning the training phase. Additionally, we faced difficulties with rendering Bangla fonts, as they were not displaying correctly, which led to errors in visual representation. Initially, our model exhibited low accuracy during the early training phases. However, after 500 epochs of training using LSTM, we overcame fluctuations in accuracy observed between epochs 310 and 330, ultimately achieving an impressive accuracy rate of 94%. Concurrently, our loss function showed a steady decline, reaching minimal values by the 500th epoch. This stability in both accuracy and loss signifies the effectiveness of the model's learning process. For a visual representation of our progress, please refer to the training accuracy and loss progression graph provided below. The successful resolution of these challenges highlights the robustness of our approach and its potential for real-world applications. Additionally, ongoing improvements to the dataset and font rendering are expected to further enhance the model's performance and usability, paving the way for broader adoption and integration in assistive technologies.

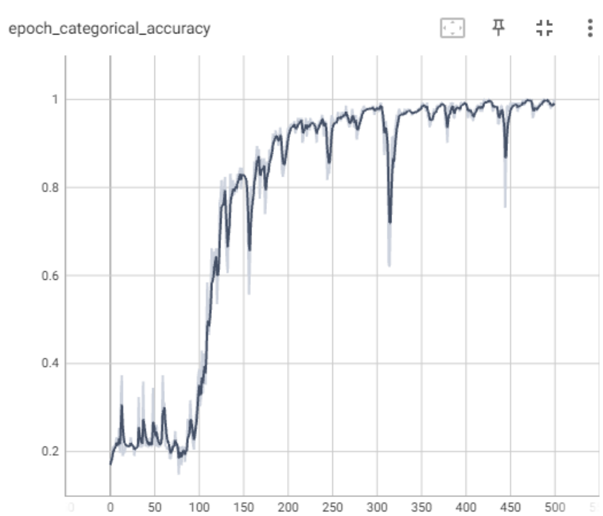

Fig. 11. Accuracy graph

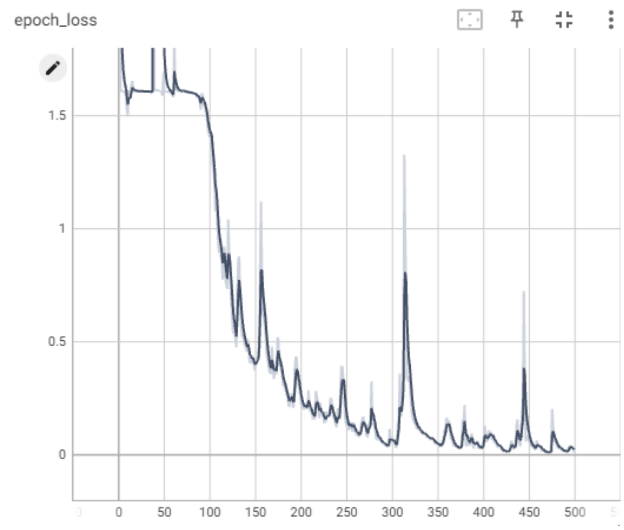

Fig. 12. Training loss graph

V. IMPACT OF BSLT

BSLT, an AI-powered innovation, is set to significantly impact third-world countries like Bangladesh, especially within the hearing-impaired community. By dismantling traditional communication barriers, BSLT aims to more fully integrate the deaf community into society. This advancement harnesses AI's transformative potential to enhance access to education, employment, and social inclusion. Moreover, BSLT represents a crucial step toward fostering equality and improving the quality of life for individuals with hearing impairments. By paving the way for greater inclusivity, it contributes to a more equitable future and empowers individuals to participate more actively in societal development.

VI. CONCLUSION

Bangla Sign language translation helps bridge communication gaps between the deaf and non-deaf communities. The integration of MediaPipe and RNN networks has empowered the development of these systems while incorporating computer vision technology takes the technology to the next level. Real-time translation enhances convenience and fluidity in communication. Additionally, the advancement of sign language translation exemplifies AI's transformative role in breaking down communication barriers, enabling more inclusive interactions, and fostering greater understanding among diverse groups. Future innovations will continue to enhance accuracy and usability, expanding the potential of this technology to reach broader.

VII. FUTURE SCOPE

Introducing sign language to voice translation represents a significant advancement, enhancing communication convenience and realism for the deaf community. This upgrade boosts confidence among deaf individuals and seamlessly integrates into daily interactions. By combining Natural Language Processing (NLP) with computer vision, we unlock the sector's true potential, enabling features like sentence auto-completion and predictive text. This innovation not only sets new standards in accessibility and inclusivity but also paves the way for groundbreaking advancements in sign language technology. Future developments will further refine these capabilities, ensuring even greater accuracy and ease of use in diverse contexts.

Additionally, integrating machine learning algorithms can facilitate continuous improvement of the translation models, adapting to various sign language dialects and user-specific nuances. Collaborative efforts with linguists and the deaf community will be crucial in training these models to recognize and interpret subtle gestures accurately. The potential for real-time translation in educational settings, workplaces, and social interactions are immense, promising to bridge communication gaps and foster a more inclusive society. By prioritizing user feedback and iterative development, we can create a robust and user-friendly translation system that significantly enhances the quality of life for deaf individuals worldwide.

ACKNOWLEDGMENT

We are immensely grateful to Dr. Shahnewaz Siddique for his unwavering support and invaluable guidance throughout our project. His expertise and commitment have been instrumental in our success. We also extend our heartfelt thanks to everyone who has contributed to this journey, offering encouragement and support that have been pivotal in achieving our goals and advancing our skills.

REFERENCES

[1] Necati Cihan Camgoz1, Simon Hadfield1, Oscar Koller2, Hermann Ney2, Richard Bowden1, 1University of Surrey, 2RWTH Aachen University. Neural Sign Language Translation. CVF 2018.

[2] jinalee Jayeshkumar Rava Electronics and Communication Department Institute of Technology, Nirma University Ahmedabad, India, Ruchi Gajjar, Elec tronics and Communication Department Institute of Technology, Nirma University Ahmedabad, India. Real-time Sign Language Recognition using Computer Vision. 2021 3rd International Conference on Signal Processing and Communication (ICPSC) — 13 – 14 May 2021 — Coimbatore.

[3] Siming He Ridley College, St. Catharines, Canada. Research of a Sign Language Translation System Based on Deep Learning. 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM).

[4] Lianyu Hu, Liqing Gao, Zekang Liu, Wei Feng College of Intelligence and Computing, Tianjin University, Tianjin 300350, China. Continuous Sign Language Recognition with Correlation Network. CVF 20233.

[5] Sign Language Translator and Gesture Recognition. Mohammed Elmahgiubi School of Engineering University of Guelph, Ontario, Canada Mohamed Ennajar National Information Security and Safety Assurance, Tripoli, Libya Nabil Drawil, and Mohamed Samir Elbuni Computer Engineering, Faculty of Engineering University of Tripoli, Tripoli, Libya.

[6] Oishee Bintey Hoque, Mohammad Imrul Jubair, Md. Saiful Islam‡, Al-Farabi Akash§, Alvin Sachie Paulson¶Department of Computer Science and Engineering,Ahsanullah University of Science and Technology Dhaka, Bangladesh,bintu3003, saiful.somum, §alfa.farabi, sachiekhan@gmail.com, †mohammadimrul.jubair@ucalgary.

[7] Bangla Sign Language Recognition using Convolutional Neural Network. Farhad Yasir1 , P.W.C. Prasad1 , Abeer Alsadoon1 , A. Elchouemi2 , Sasikumaran Sreedharan3 1 Charles Sturt University Study Centre, Sydney, Australia 2 Walden University, USA 3 Department of Computer Engineering, College of Computer Science, King Khalid University, KSA.

[8] Hierarchical LSTM for Sign Language Translation. Dan Guo,1 Wengang Zhou,2 Houqiang Li,2 Meng Wang1 1School of Computer and Information, Hefei University of Technology, Hefei, 230009, China 2EEIS Department, University of Science and Technology of China, Hefei, 230027, China.

[9] Sign Language Recognition using LSTM and Media Pipe. G. Mallikarjuna Rao Department of computer science and engineering Gokaraju Rangaraju Institute Of Engineering and Technology Hyderabad, India P. A.Sujasri Department of computer science and engineering Gokaraju Rangaraju Institute Of Engineering and Technology Hyderabad, India Cheguri Sowmya Department of computer science and engineering Gokaraju Rangaraju Institute Of Engineering and Technology Hyderabad, India Soma Anitha Department of computer science and engineering Gokaraju Rangaraju Institute Of Engineering and Technology Hyderabad, India Dharavath Mamatha Department of computer science and engineering Gokaraju Rangaraju Institute Of Engineering and Technology Hyderabad, India Ramavath Alivela Department of computer science and engineering Gokaraju Rangaraju Institute Of Engineering and Technology Hyderabad, India.

[10] Sign language Recognition Application Using LSTM and GRU. Pranav Sheth Indus University, Sanju Rajora Indus University, April 2023.

[11] Tutorial on LSTMs: A Computational perspective. Manu Rastogi.

[12] Neural network and Deep learning. By MichaelNielson Dec 2019.

[13] Solaiman Lipi Bangla font.

[14] Ishara-Lipi: The First Complete MultipurposeOpen Access Dataset of Isolated Characters for Bangla Sign Language. Md. Sanzidul Islam Daffodil International University, Akm Shahariar Azad Rabby University of Alabama at Birmingham, Sadia Sultana Sharmin Daffodil International University

[15] Jiang, S.; Sun, B.; Wang, L.; Bai, Y.; Li, K.; Fu, Y. Skeleton aware multi-modal sign language recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 21–24 June 2021.

[16] De Coster, M.; Herreweghe, M.V.; Dambre, J. Sign Language Recognition with Transformer Networks. In Proceedings of the Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 13–15 May 2020 .

[17] M. Melnyk, V. Shadrova, and B. Karwatsky, "Article: Towards computerassisted international sign language recognition system: A systematicsurvey," International Journal of Computer Applications, March 2014, published by Foundation of Computer Science,New York, USA.

[18] . Wang, W. Gao, and Z. Xuan, "A real-time large vocabulary continuous recognition system for chinese sign language," in Advances inMultimedia Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg,2001.

[19] Aditya Das, Shantanu Gawde, Khyati Suratwala and Dr. Dhananjay Kalbande. "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images".In: International Conference on Smart City and Emerging Technology (ICSCET), 2018 .

[20] F. M. Rahim, T. E. Mursalin, and N. Sultana, "Intelligent sign language verification system–using image processing, clustering and neural network concepts," International Journal of Engineering Computer Science and Mathematics, 2010.

[21] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning.MIT press, 2016.

[22] E.-J. Ong, H. Cooper, N. Pugeault, and R. Bowden. Sign Language Recognition using Sequential Pattern Trees. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[23] P. Buehler, A. Zisserman, and M. Everingham. LearningSign Language by Watching TV (using weakly aligned subtitles). In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[24] H. Liu, and C. Zhang. Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017

[25] X. Glorot and Y. Bengio. Understanding the Difficulty ofTraining Deep Feedforward Neural Networks. In International Conference on Artificial Intelligence and Statistics,2010.

[26] A. Graves, A.-r. Mohamed, and G. Hinton. Speech Recognition with Deep Recurrent Neural Networks. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[27] . Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. Show, Attend and Tell: NeuralImage Caption Generation

with Visual Attention. In International Conference on Machine Learning (ICML), 2015.

[28] -D. Yang, S. Sclaroff, and S.-W. Lee. Sign Language Spotting with a Threshold Model based on Conditional Random Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 31(7), 2009.

[29] F. Yin, X. Chai, and X. Chen. Iterative Reference Driven Metric Learning for Signer Independent Isolated Sign Language Recognition. In European Conference on ComputerVision (ECCV), 2016.

[30] Likhar, P. Bhagat, N.K. Rathna, G.N. Deep Learning Methods for Indian Sign Language Recognition. In Proceedings of the 2020 IEEE 10th International Conference on Consumer Electronics,Germany.

[31] Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. Neural Comput.

[32] Chen, J. CS231A Course Project Final Report Sign Language Recognition with Unsupervised Feature Learning. 2012

[33] Agarwal, S.R.; Agrawal, S.B.; Latif, A.M. Article: Sentence Formation in NLP Engine on the Basis of Indian Sign Language using Hand Gestures. Int. J. Comput. Appl. 2015

[34] REN S,HE K,GIRSHICK R,et al. Faster r -cnn: To-wards real - time object detection with region proposal net-works[J]. Advances in neural information processing sys-tems,2015

[35] Ronglai Zuo and Brian Mak. Consistency-enhanced continuous sign language recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition

[36] Victor Wiley1, Thomas Lucas2 Teknik Informatika, TSCC Jakarta Indonesia1 victorwiley10@gmail.com,Computer Vision and Image Processing: A Paper Review.International Journal Of Artificial Intelegence Research ISSN: 2579-7298 Vol 2, No 1, June 2018, pp. 28-36.

[37] M. Hrúz, J. Trojanová, M. Železný. (2012). Local Binary Pattern based features for sign language recognition. Pattern Recognition and Image Analysis.