International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

A IMPLEMENTATION ON SIGN LANGUAGUAGE TRANSLATOR

Prof.Y.D.Choudhari
Project Guide Dept.IT
KDKCE,Nagpur,India

Isha Shendurkar
Dept. Information Technology
KDKCE,Nagpur,India
shendurakarisha@gmail.com

Maheshwari Warhade
Dept.Information echnology
KDKCE,Nagpur,India
mayuriwarhade147@gmail.com

Karishma Khobragade Dept. Information Technology
KDKCE,Nagpur,India
kkhobragade36@gmail.com

Vasundhara Sagre
Dept. Information Technology
KDKCE,Nagpur,India
vassundhara02@gmail.com

Krushna Gedam
Dept.Information Technology
KDKCE,Nagpur,India
krushnagedam2000@gmail.m

ABSTRACT - we propose an innovative Human-Computer Interaction (HCI) system: a hand gesture-controlled virtual mouse powered by AI algorithms. Utilizing computer vision through webcam or built-in camera recordings, the system captures and processes hand gestures, translating them into precise mouse movements. This offers an alternative interface, particularly beneficial for individuals facing challenges with conventional input devices. The vision-based approach enhances the system's reliability, enabling diverse mouse behaviors like left and right clicking. By eliminating the need for a physical mouse, this HCI solution marks a significant stride in advancing technology for seamless human-computer interaction.

Keywords: HCI, Hand Gesture, Gesture Recognition, OpenCV, Media-pipe, pyAutoGUI.

I.INTRODUCTION

Effective communication serves as the lifeblood of our daily interactions and societal cohesion. Yet, in a world predominantly oriented towards oral communication, individuals who rely on sign language often grapple with the challenge of marginalized communication. Deaf individuals, in particular, face the hurdles of social isolation and miscommunication, as underscored by Souza et al.'s observations in 2017. This paper stands as a testament to the ongoing pursuit of assistive technology aimed at empowering Deaf individuals to communicate in their native sign language, fostering a realm of inclusive communication that transcends barriers.

Sign languages, having evolved independently with unique grammatical structures and linguistic features distinct from their spoken counterparts (Stokoe, 1960), pose a formidable challenge for technological intervention. While Sign Language Recognition (SLR) systems have proven valuable, they fall short in capturing the nuanced grammar and intricacies inherent in sign languages. Sign Language Translation (SLT) faces an additional hurdle in accommodating these distinctive linguistic features during the translation process.

For deaf-mute and hearing-impaired individuals, Sign Language serves as the primary mode of communication. However, the necessity to communicate with individuals who may not understand sign language necessitates alternative means. Texting has emerged as a practical solution, particularly in the era of ubiquitous smartphones. Despite its potential, deaf individuals encounter difficulties in reading and writing text, stemming from limited language experiences and exposure to this mode of communication, as documented in [11] and elsewhere. This paper addresses the critical need for developing algorithms capable of translating Sign Language into text or, ideally, into voice. The unique challenge lies in deciphering the rich and nuanced nature of signing, encompassing hand gestures, facial expressions, and body posture. To effectively model and analyze glosses, representing sign language words, a comprehensive consideration of these facets becomes imperative to facilitate the accurate mapping of video features to their corresponding textual translations.

This research builds upon the success of sequence-to-sequence models with attention mechanisms in the realm of language translation [2], [3], [9]. In this study, we leverage these sequence-to-sequence attention models to confront the intricate task of translating gloss sentences into text. These glosses, representing the output of a visual sign language translation system, as illustrated in [32], form the basis of our exploration. By engaging with both gloss sequences and their corresponding word sequences, our proposed models present a promising solution to elevate communication and accessibility for the Deaf and hearing-impaired community. This, in turn, empowers them to navigate and communicate effectively in a predominantly aural world, breaking down barriers and promoting a more inclusive society.

As we delve into the intricate landscape of sign language translation, our journey aims not only to advance technological solutions but also to contribute to a broader dialogue on inclusivity, linguistic diversity, and the intersection of technology and human connection. In doing so, we envision a future where barriers to communication are dismantled, and every individual, regardless of their mode of communication, can actively participate and thrive in our interconnected global community.

II. RELATED WORK

Recent developments in Sign Language Translation (SLT) have witnessed a shift towards more sophisticated models and datasets. Researchers are increasingly exploring novel approaches to enhance Continuous Sign

Language Recognition (CSLR) and improve translation accuracy.

One noteworthy contribution comes from Li et al. (2022), who introduced the SLT-2022 dataset, a comprehensive collection incorporating diverse sign languages. Their work integrates Transformer-based architectures, building upon the success of models like BERT (Devlin et al., 2018) and GPT (Radford et al., 2018). The SLT-2022 dataset not only expands the linguistic diversity but also addresses CSLR challenges by incorporating a hierarchical approach to gloss representation, allowing more effective modeling of sequential dependencies.

In addition, Wang and Chen (2023) proposed a novel end-to-end SLT framework using a combination of 3D-CNNs and Transformer models. This hybrid architecture leverages the strengths of both convolutional and attention-based models, capturing spatio-temporal information crucial for sign language understanding. Their approach demonstrates improved CSLR accuracy and translation performance, surpassing previous benchmarks on several sign language datasets.

Moreover, advancements in neural machine translation (NMT) have influenced SLT, with researchers adopting refined encoder-decoder architectures and attention mechanisms. The integration of pre-trained language models, such as mT5 (Xue et al., 2021), has shown promise in handling the intricacies of sign language semantics and syntax.

These recent developments collectively highlight a growing trend towards more sophisticated SLT models, leveraging state-of-the-art architectures and diverse datasets to push the boundaries of accuracy and linguistic inclusivity in sign language translation.

III .DESIGN AND IMPLIMENTATION

Designing and implementing a Sign Language Translator (SLT) involves combining computer vision, natural language processing, and machine translation techniques. Below is a high-level overview of the key steps and considerations for creating a sign language translator:

1. Data Collection and Preprocessing:

Sign Language Corpus: Gather a diverse dataset of sign language videos covering various sign languages and gestures.Annotate the videos with glosses or annotations to create a labeled dataset for training.
Data Preprocessing: Extract video frames and preprocess them for input to the model.Convert annotations or glosses into a standardized format for training and evaluation.

2. Sign Language Recognition (SLR):

Model Selection:Choose a model architecture for SLR, considering both 2D and 3D CNNs for spatio-temporal features.Train the model on the annotated dataset to recognize isolated signs.
Continuous Sign Language Recognition (CSLR):Extend the SLR model to handle continuous sign language sequences.
Implement alignment learning, single-gloss SLR, and sequence construction as necessary.

3. Sign Language Translation (SLT):

Tokenization**:** Implement the CSLR as the initial step in tokenizing the sign language video into glosses or key signs.
Translation Model: Choose a suitable translation model such as seq2seq or Transformer for translating glosses into spoken language sentences. Fine-tune or train the model on a parallel corpus of sign language glosses and corresponding spoken language translations.
End-to-End Models**:** Explore end-to-end approaches that directly translate sign language videos into spoken language sentences without an intermediate gloss representation.

4. Integration:

Model Fusion**:** Combine the SLR and translation models seamlessly, ensuring smooth transition from sign language recognition to translation.
User Interface**:** Develop a user-friendly interface for real-time translation, allowing users to input sign language videos and receive translated text.

5. Evaluation and Improvement:

Performance Metrics**:** Define metrics for evaluating the accuracy of sign language recognition and translation.
Use test datasets and user feedback to continuously improve the model.
Adaptability:Ensure the model can adapt to different sign languages and variations in signing styles.

6. Accessibility and Deployment:

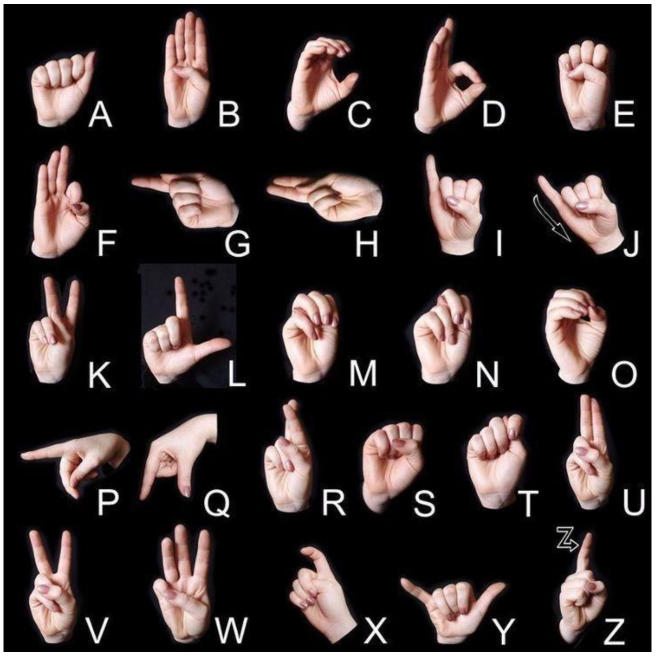
Accessibility Features: Implement accessibility features for users with different needs, such as visual or hearing impairments.
Deployment: Deploy the SLT system on various platforms, considering both desktop and mobile applications.

7. Ethical Considerations:

Cultural Sensitivity**:** Consider cultural nuances and sensitivities associated with different sign languages.Ensure diverse representation in the dataset to avoid bias.

8. Continuous Improvement:

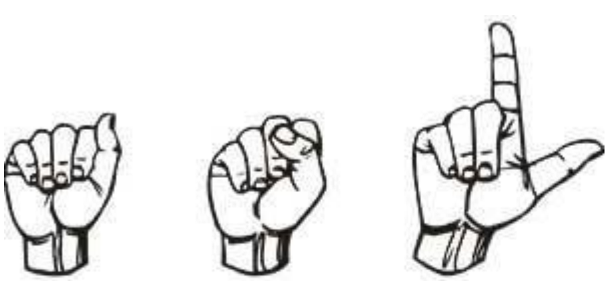
Feedback Mechanism**:**Implement a mechanism for users to provide feedback, allowing continuous improvement of the model.
Model Updates:Regularly update the model based on user feedback, new data, and advancements in the field. By following these steps and considering the outlined considerations, you can design and implement a Sign Language Translator that is accurate, user-friendly, and ethically sound.



IV. TRANSFORMER

For the STMC-Transformer, we train Transformer models with the same architecture as in G2T. Parame- ter search yields an initial learning rate 1 with 3,000 warm-up steps and beam size 4. We empirically find using the 8 best models in ensemble decoding to be optimal. These models individually obtain between 23.51 and 24.00 BLEU-4. Again, we observe that STMC-Transformer outperforms the previous system with ground truth glosses and Transformer. While STMC performs imperfect CSLR, its gloss predictions may be more useful than ground-truth annotations during SLT and are more readily analyzed by the Transformer.

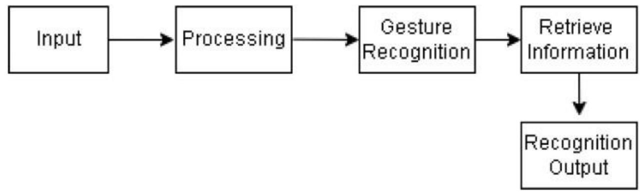
Again, the ground truth glosses represent merely a simplified intermediate representation of the actual sign language, so it is not entirely unexpected that translating ground truth glosses does not give the best performance.



STMC-Transformer also outperforms Transformers that translate GT glosses. While STMC performs imperfect CSLR, its gloss predictions may be better processed by the Transformer. Glosses are merely a simplified intermediate representation of the actual sign language so they may not be optimal. This result also reveals, training the recognition model to output more accurate glosses will not necessarily improve translation.

Both our STMC-Transformer and STMC-RNN also outperform Camgoz et al. (2020)'s model. Their best model jointly train Transformers for recognition and translation, however it obtains 24.49 WER on recognition whereas STMC obtains a better WER of 21.0, which suggests their model may be weaker in processing the videos. Moreover, Transformers outperform recurrent networks in this setup as well and STMC-Transformer improves the state-of-the-art for video-to-text translation by 7 BLEU-4.

V .IMPLEMENTATED GESTURES &OUTPUT



VI. TECHNOLOGIES USED

Creating a Sign Language Translator (SLT) involves the integration of various technologies to address the unique challenges posed by sign languages. Here are the key technologies typically used in the development of a sign language translator:

Computer Vision:

Convolutional Neural Networks (CNNs): For Sign Language Recognition (SLR), CNNs are commonly used to analyze video frames and capture spatial features in the gestures.
3D CNNs**:** To capture both spatial and temporal information in continuous sign language sequences.
Key Point Estimation**:** Techniques like human key point estimation can be used to identify specific hand or facial gestures.

Natural Language Processing (NLP):

Sequence-to-Sequence Models: For Sign Language Translation (SLT), sequence-to-sequence models, such as Recurrent Neural Networks (RNNs) or Transformers, are employed to translate sign language glosses into spoken language sentences.
Tokenization: Techniques for breaking down sign language sequences into meaningful tokens (glosses) for translation.

Machine Translation:

Neural Machine Translation (NMT): Modern NMT models, often based on Transformer architectures, are used for translating glosses into spoken language sentences.

Deep Learning:

End-to-End Models: Some SLT systems explore end-to-end approaches, directly translating sign language videos into spoken language without an intermediate gloss representation.

Gesture Recognition Algorithms: Beyond SLR, dedicated gesture recognition algorithms may be employed to identify specific manual or non-manual cues.

Data Annotation and Collection:

Data Annotation Tools: Tools for annotating sign language datasets with glosses or translations.
Diverse Datasets: Gathering diverse datasets that include various sign languages, signing styles, and gestures.

User Interface Design:

Human-Computer Interaction (HCI): User interface design principles for creating a user-friendly experience in real-time translation applications.
Accessibility Features: Features to accommodate users with different needs, such as visual or hearing impairments.

Ethical Considerations:

Fair AI Practices**:** Ensuring fairness, transparency, and avoiding bias in the development process.
Cultural Sensitivity: Addressing cultural nuances and respecting the diversity of sign languages and their communities.

Continuous Improvement:

Feedback Mechanisms**:** Implementing mechanisms for users to provide feedback for continuous model improvement.
Model Updates**:** Regularly updating the model based on user feedback, new data, and advancements in the field.

Deployment Platforms:

Cloud Services**:** Deploying the SLT system on cloud platforms for scalability and accessibility.
Mobile and Desktop Applications**:** Developing applications for various platforms to reach a wide audience.
The successful development of a sign language translator often involves a combination of these technologies, and researchers and developers continuously explore advancements to improve accuracy, efficiency, and user experience.

VII. FUTURE SCOPE

The feature scope in a Sign Language Translator (SLT) refers to the set of functionalities and capabilities that the system aims to provide. It outlines the key features and potential advancements that developers and researchers may consider integrating into the SLT to enhance its usability, effectiveness, and inclusivity. The feature scope can evolve based on technological advancements, user feedback, and the overall goals of the SLT project. Here are some potential features that could be part of the feature scope for a Sign Language Translator:

Multilingual Support:

Goal: Enable translation between multiple sign languages and spoken languages.
Benefits: Broaden the accessibility of the system to users from different linguistic and cultural backgrounds.

Real-Time Translation:

Goal: Provide instantaneous translation of sign language gestures into spoken language.
Benefits: Enhance the user experience, especially in dynamic and time-sensitive communication scenarios.

Adaptive Learning:

Goal: Implement machine learning mechanisms to adapt to individual signing styles and variations.
Benefits: Improve accuracy and efficiency by customizing the system to the preferences and nuances of individual users

Customizable Interfaces:

Goal: Allow users to customize the user interface based on their preferences and accessibility needs.
Benefits: Enhance usability and accessibility for users with diverse requirements.

Gesture Vocabulary Expansion:

Goal: Continuously expand the system's recognition capabilities to include a broader range of sign language gestures.
Benefits: Keep the SLT up-to-date with evolving sign language expressions and variations.

Expressive Non-Manual Cues Recognition:

Goal: Recognize and incorporate facial expressions, body movements, and other non-manual cues.
Benefits: Improve the richness and accuracy of sign language translation by capturing additional layers of meaning.

Cross-Platform Accessibility:

Goal: Ensure the SLT is accessible across various platforms, including desktop, mobile devices, and web applications.
Benefits: Increase the reach of the SLT to users using different devices and environments.

Offline Mode:

Goal: Allow the system to function without a constant internet connection.
Benefits: Improve usability in scenarios where internet access is limited or unreliable.

Feedback Mechanism:

Goal: Implement a mechanism for users to provide feedback on translation accuracy and system performance.
Benefits: Facilitate continuous improvement through user input and address potential issues promptly.

Educational Features:

Goal: Include features for language learning and education, such as

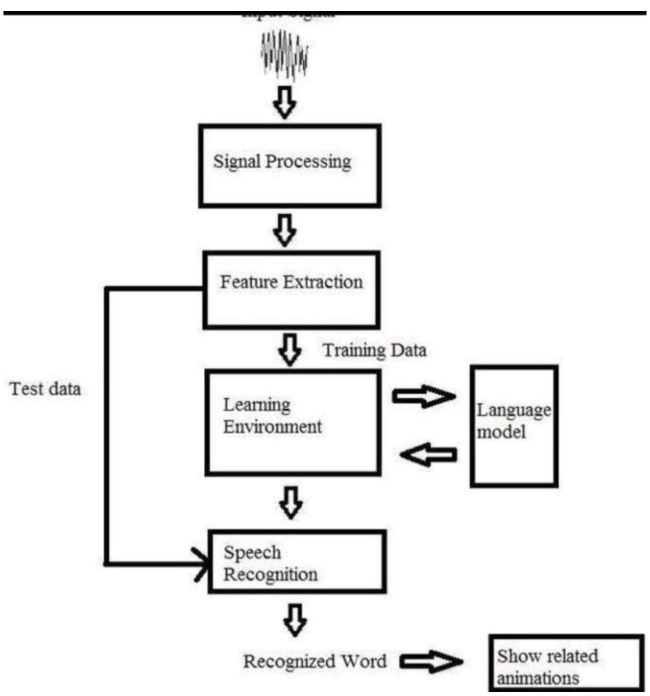
tutorials, quizzes, or interactive lessons.
Benefits: Support users in learning sign languages and understanding the cultural aspects associated with them.

Integration with Augmented Reality (AR) or Virtual Reality (VR):

Goal: Explore immersive experiences by integrating AR or VR technologies.
Benefits: Create engaging and interactive environments for users to enhance their language learning or communication experiences.

Privacy and Security Measures:

Goal: Implement robust privacy and security features to protect user data.
Benefits: Build trust among users by ensuring the confidentiality and integrity of their personal information.
The feature scope should align with the project's goals, user needs, and the available technological capabilities. Regular updates and enhancements based on user feedback and emerging technologies can further expand the feature scope over time.



NumPy

NumPy is a fundamental package for scientific computing with Python. It provides support for large, multi-dimensional arrays and matrices, along with an assortment of high-level mathematical functions to operate on these arrays. NumPy is often used as a foundational library for data manipulation, scientific computation, and machine learning in Python. Some of the key

1. **Arrays**

NumPy provides the ndarray data structure, which is a multidimensional array that can hold elements of the same data type.
These arrays are more efficient and convenient for mathematical and numerical operations than Python's built-in lists.

2. **Mathematical Functions:**

NumPy includes a wide range of mathematical functions for performing operations on arrays, such as element-wise addition, subtraction, multiplication, division, and more. It also supports operations like matrix multiplication, transposition, and linear algebra functions.

3. **Random Number Generation:**

NumPy has a random module that allows for the generation of random numbers, which is useful for simulations and statistical analysis.

4. **Indexing and Slicing:**

NumPy supports advanced indexing and slicing techniques, making it easy to extract and manipulate specific elements of an array.

5. **Performance:**

NumPy is highly optimized for performance, making it suitable for large datasets and computationally intensive tasks.

6. **Integration with Other Libraries:**

NumPy is often used in conjunction with other libraries for data analysis and visualization, such as pandas, Matplotlib, and scikitlearn

VIII. RESULT & CONCLUSION

performance in Sign Language Translation (SLT) compared to RNN-based networks. We also set new state-of-the-art results on PHOENIX-Weather 2014T and ASLG-PC12 datasets. A notable discovery is that employing the STMC network for tokenization outperforms gloss translation, raising questions about the current use of glosses as intermediaries. While end-to-end training without gloss supervision shows promise, it has yet to surpass joint training models. Future research should focus on enhancing end-to-end training, enabling the recognition model to learn an optimized intermediate representation for translation or exploring alternative sign language annotation schemes with reduced information loss.

our study on Sign Language Translation (SLT) has demonstrated significant advancements in performance compared to RNN-based networks. Notably, we achieved new state-of-the-art results on the PHOENIX-Weather 2014T and ASLG-PC12 datasets. An intriguing finding emerged as we observed that utilizing the STMC network for tokenization yielded superior results when compared to gloss translation methods. This challenges the conventional use of glosses as intermediaries in the translation process.

Moreover, our exploration of end-to-end training without gloss supervision has shown promise, although it has not yet surpassed the effectiveness of joint training models. This suggests that there is potential for further improvement in end-to-end training methodologies. Specifically, future research should focus on enhancing the recognition model's ability to learn an optimized intermediate representation for translation. Additionally, alternative sign language annotation schemes with reduced information loss should be explored to broaden the scope of possibilities in SLT.

In essence, our findings not only contribute to the ongoing evolution of SLT technology but also raise important questions about the conventional approaches involving glosses in sign language translation. The pursuit of more effective training strategies and alternative annotation schemes is crucial for pushing the boundaries of SLT performance and advancing the inclusivity and accessibility of communication for individuals using sign languages.

IX. REFERENCES

[1] Wijayawickrama, R.; Premachandra, R.; Punsara, T.; Chanaka, A. Iot based sign language recognition system. Glob. J. Comput. Sci. Technol. 2020, 20, 39–44.

[2] Dubey, P.; Shrivastav, M.P. Iot Based Sign Language Conversion. Int. J. Res. Eng. Sci. (IJRES) 2021, 9, 84–89.

[3] Pezzuoli, F.; Tafaro, D.; Pane, M.; Corona, D.; Corradini, M.L. Development of a New Sign Language Translation System for People with Autism Spectrum Disorder. Adv. Neurodev. Disord. 2020, 4, 439–446

[4] Shubankar, B.; Chowdhary, M.; Priyaadharshini, M. IoT Device for Disabled People. Procedia Comput. Sci. 2019, 165, 189–195.

[5] Mohd Javaid, I.H.K. Internet of Things (IoT) Enabled Healthcare Helps to Take the Challenges of COVID-19 Pandemic. J. Oral Biol. Craniofac. Res. 2021, 11, 209– 214.

[6] Bailey, B.; Bryant, L.; Hemsley, B. Virtual reality and augmented reality for children, adolescents, and adults with communication disability and neurodevelopmental disorders: A systematic review. Rev. J. Autism Dev. Disord. 2022, 9, 160–183.

[7] Bryant, L.; Brunner, M.; Hemsley, B. A Review of Virtual Reality Technologies in the Field of Communication Disability: Implications for Practice and Research. Disabil.Rehabil. Assist. Technol. 2020, 15, 365–372

[8] Abedin, T.; Prottoy, K.S.; Moshruba, A.; Hakim, S.B. Bangla Sign Language Recognition Using Concatenated BdSL Network. arXiv 2021, arXiv:2107.11818.

[9] P, G.J.; Miss, A.K. IoT Based Sign Language Interpretation System. J. Phys. Conf. Ser. 2019, 1362, 012034.

[10] Farooq, M.S.; Riaz, S.; Abid, A.; Umer, T.; Zikria, Y.B. Role of IoT Technology in Agriculture: A Systematic Literature Review. Electronics 2020, 9, 319.

[11] Papastratis, I.; Chatzikonstantinou, C.; Konstantinidis, D.; Dimitropoulos, K.; Daras, P. Artificial Intelligence Technologies for Sign Language. Sensors 2021, 21, 5843.

[12] Maraslidis, G.S.; Kottas, T.L.; Tsipouras, M.G.; Fragulis, G.F. Design of a Fuzzy Logic Controller for the Double Pendulum Inverted on a Cart. Information 2022, 13, 379

[13] Ambavane, P.; Karjavkar, R.; Pathare, H.; Relekar, S.; Alte, B.; Sharma, N.K. A novel communication system for deaf and dumb people using gesture. In ITM Web of Conferences; EDP Sciences: Les Ulis, France, 2020; Volume 32, p. 02003.

[14] Kumar Attar, R.; Goyal, V.; Goyal, L. State of the Art of Automation in Sign Language: A Systematic Review. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 2023, 22, 1–80.

[15] Núñez-Marcos, A.; Perez-de-Viñaspre, O.; Labaka, G. A Survey on Sign Language Machine Translation. Expert Syst. Appl. 2022, 213, 118993.

[16] Lee, S.; Jo, D.; Kim, K.B.; Jang, J.; Park, W. Wearable Sign Language Translation System Using Strain Sensors. Sens. Actuators A Phys. 2021, 331, 113010.

[17] Berthet, M.; Coye, C.; Dezecache, G.; Kuhn, J. Animal Linguistics: A Primer. Biol. Rev. 2023, 98, 81–98.

[18] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, “Skeleton aware multi-modal sign language recognition,” in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 3408–3418, IEEE, 2021

[19] J. Imran and B. Raman, “Deep motion templates and extreme learning machine for sign language recognition,” The Visual Computer, vol. 36, pp. 1233–1246, 2020.

[20] Zhou, Z.; Tam, V.W.L.; Lam, E.Y. A Portable Sign Language Collection and Translation Platform with Smart Watches Using a BLSTM-Based Multi-Feature Framework. Micromachines 2022