Sign Language Generation and Translation for Hearing Disabled People Using Resnet and Tts Engine

Indira Chetty B
B.Tech., Artificial intelligence and data science
Sri Krishna College of Engineering and Technology
Coimbatore, India

Muthukumaran L
B.Tech., Artificial intelligence and data science
Sri Krishna College of Engineering and Technology
Coimbatore, India

Hari surya M R
B.Tech., Artificial intelligence and data science
Sri Krishna College of Engineering and Technology
Coimbatore, India

***Abstract*— A dual purpose device for individuals with sensory disabilities, seamlessly combining sign language generation and translation. By leveraging laptop vision technologies, the machine offers real-time interpretation of signal gestures. This challenge makes use of superior deep getting to know techniques, together with ResNet, to decorate gesture interpretation for progressed translation accuracy and flexibility to numerous signing patterns. A key function is the incorporation of deep nearest neighbor detection for unique hand reputation, making sure of accurate interpretation of problematic gestures. Users can also personalize the gadget by schooling it with their own gestures, developing a greater user-centric revel in. To sell inclusivity, the machine consists of a real-time video name option with the use of the TokBox API, enhancing conversation in the signal language context. This progressive method addresses demanding situations faced by using the disability community, consisting of the dearth of available real-time translation tools and signal language interpreters. By combining current technologies, the mission objectives are to beautify translation accuracy, improve hand reputation, and offer effective communication guides for users with sensory disabilities, in the long run fostering inclusivity and accessibility.**

Keywords— Speech Recognition, Natural Language Processing, Machine Learning Models, Deep Learning, Feature Extraction, Multilingual Support, Real-time Adaptability, Contextual Understanding, User Interface Design, Adaptive Learning Mechanisms, Data Flow Testing, Augmented Reality, Virtual Reality, Edge Computing, Accessibility Technology)

I. INTRODUCTION

This project aims to redefine signal language interaction by means of breaking down conversation limitations through real-time translation and numerous style reputation. While signal language is a crucial mode of conversation for lots individuals, existing solutions regularly battle with inaccuracies and obstacles in adapting to numerous signing patterns and vocabulary. To address those demanding situations, the undertaking combines signal language generation and translation, making use of laptop vision technology for real-time interpretation of sign gestures. It introduces an innovative sign language translation gadget that employs ResNet and advanced deep studying techniques for unique gesture interpretation. Users are empowered to train the device on their particular gestures and expand a custom signal library, ensuring accurate translation and facilitating nuanced verbal exchange despite non-general signing patterns and private preferences. Leveraging advanced technology along with ResNet18 and deep mastering strategies, the undertaking enhances translation accuracy and comprises diverse signing patterns. Additionally, the gadget consists of deep nearest neighbor detection for delicate hand recognition, allowing precise interpretation of complicated gestures. To beautify inclusivity, the machine integrates a real-time video call choice the usage of the TokBox API, enriching the user revel in and promoting seamless communication in the sign language context. Overall, this mission represents a good-sized advancement in enhancing accessibility and communication for individuals with sensory disabilities, empowering them to communicate more successfully and independently in their everyday lives.

RELATED WORK

[1]SIGNFORMER: Deep Vision Transformer for Sign Language Recognition by Chintan M. Bhatt, Tanzila Saba (2022)

Signformer presents a groundbreaking vision-based transformer network specifically designed for Sign Language Recognition (SLR). Harnessing advanced deep learning approaches and transformer architectures, the system is poised to achieve outstanding performance in discerning and understanding sign language gestures. Through the integration of vision and transformer models, Signformer elevates the precision and timeliness of real-time sign language recognition, holding the potential to revolutionize communication accessibility.

DISADVANTAGE: Also, the fact that currently existing signs communicating with sign language might be in one particular style for quite a long time and might not understand many gestures and signs which will make the communication difficult. The SVM represents a resorting to the well-established machine learning techniques that most likely or not can not seize from the new discovery given by the latest state-of-art tool of deep learning. The identification of the volatility gradient can prove to be one of the several aspects where it is problematic to obtain speed optimization which leads to the reduction in performance figure.

[2]Indian Sign Language Recognition System Using SURF with SVM and CNN (2022)

The device is designed to understand Indian Sign Language (ISL) gestures, achieving an accuracy of 84% with the aid of using SVM. The take a look at emphasizes the importance of dataset first-rate and range, displaying SVM's ability in ISL recognition. Nevertheless, the challenge acknowledges limits related to dataset length and range, that could affect the gadget's generalization to unseen information.

DISADVANTAGE: In the first place, data size and samples ones that are different have been the most important aspect of the training not just during the training phase but also later on during the implementation of software. On that subtopic, the significance of amassing more complex and more diversified data sets to augment the accuracy of the system is further addressed. From the big problems of the system's accuracy when used with gestures that the system has been studied to a small issue of the potential applicability of the rules for the gestures different from those that the system has been trained on.

[3]DeepASLR: Convolutional neural network based human-computer interface for American Sign Language recognition for hearing-impaired individuals (2022)

DeepASLR is a new challenge based on Convolutional Neural Network (CNN) for human-computer interface targeting American Sign Language (ASL) recognition in hearing impaired people The deep learning software has been enhancing the interpretation and understanding of ASL gestures, It offers a positive way of enabling communication between deaf persons, Produces an efficient and reliable way of teaching ASL into electronic format thus improving accessibility among hearing-impaired users.

DISADVANTAGE: Fundamental to deep learning since DeepASLR is what it is and has its pluses and minuses. The second issue, which might be the more problematic one would be to get a decent size and comprehensive dataset as a prerequisite for the system to be able to differentiate the ASL gesture and make the results more precise. The well-defined data loss is the main factor which hampers the systems' generalized efficiency of various signed languages' styles and expressions. After that the vast majority of the work has been done using cnn they will need even extraordinary amounts of processing power and this can be a problem in terms of getting models for real time use or in such devices that have rather limited power. Moreover, besides the mentioned issues, which are vital in order provide the best productivity level and the eased usability, also require very much the engagement from a side of a project group.

[4] A Computer Vision-Based Hand Gesture Recognition for Human-Robot Interaction: A Review (2021)

This project summarizes the work of computer vision-based hand gesture recognition systems, which are used for humanization-integrated robotic systems. It analyzes the progresses, methods and implementations in this area, considering the same as utilizing vision based techniques for right and responsive recognizing of gestures. The review is established with a purpose of providing insights from the current-technological and system-based approaches used in the communication between humans and robots with hand gestures.

DISADVANTAGE: The introduction of computer vision-based hand gestures does not mean they are perfect, since there are many challenges ahead. First and foremost there is always a risk of things like appearance changes such as brightness, lamps, obstructions or distinct background noise. These impediments may cause recognition system to lose its accuracy and reliability severly, and therefore result in the drop in application performance level. Remediating these hurdles prove essential in the development of sound and reliable gesture recognition in human robot interaction.

MATERIALS AND METHODS

The proposed architecture used various methods for sign language generation and translation.

A. *HAND RECOGNITION USING DEEP NEAREST NEIGHBOR DETECTION*

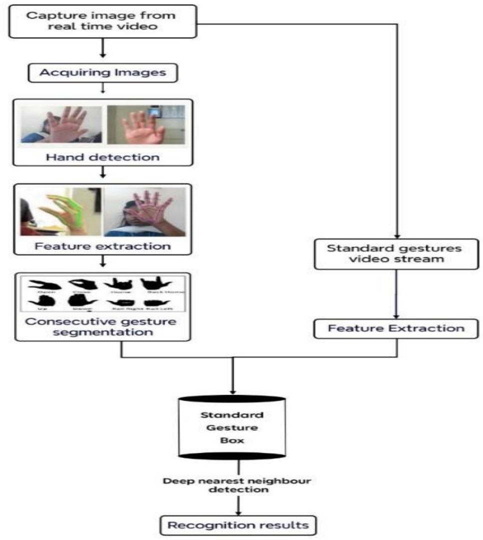


Figure 1. Hand recognition using DNN

The Process of taking pictures photographs from actual-time video involves several key steps every contributing to the successful identification and isolation of desired frames. This document describes the technical elements of this method, offering a expert explanation that would be suitable for documentation purposes.

1. Acquiring:

The first step is to get frames from a continuous video. This typically includes hardware such as a camera or software such as a webcam, which capture images at an agreed rate (e.g. frames per second) dictated by application requirements

2. Hand Detection and Feature Extraction:

Once the images are flowing, the next step is to identify relevant segments of interest. It generally uses hand recognition algorithms. These algorithms use techniques such as machine learning or image processing to analyze individual frames to determine if alcohol is present in the video stream. When the hand is identified, items will be extracted from the identified areas; These factors include: hand size, shape, movement, and relative position in the frame

3. Gesture Recognition:

In the long run, following identification of different gestures that may be recognized; the gesture type undergoes a final processing by the application based on its intended purposes. This can involve converting the gesture into actionable commands that control devices, initiating events inside the system, or giving feedback to the user. The "recognition results" phase comes out as the definitive output of this stage with respect to what was interpreted from real-time video stream in terms of meaning extracted from captured image.

4. Consecutive Gesture Segmentation:

Sequential movements of the hands involving multiple gestures may require a sequential division of gestures in some cases. At this point, we analyze the video stream surrounding the identified gestures in order to isolate individual gestures. Temporal segmentation algorithms typically use changes in hand components over time to classify them into individual gestures.

5. Feature Extraction and Recognition:

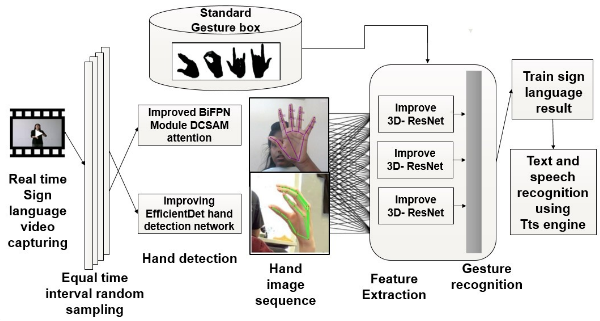
Just as in the original gesture recognition process, for each segmented gesture relevant features are extracted and fed into an identification system. Depending on the complexity of an application, this stage can involve detection of specific variability within a recognized gesture

6. Processing and Recognition Results:

At length, gestures that can be seen after sight; the gesture type is the final operation performed by the application based on its purpose. This signal may involve modifying executable commands that control devices, initiating events in the system, or providing information to the user. The "visual results" phase emerges as a definitive result of this phase in terms of what was interpreted in terms of the meaning extracted from the captured footage from the real-time video flow.

B. *SIGN LANGUAGE TRANSLATION USING RESNET*

A new approach to hand gesture recognition is proposed with regard to improving both interpretation of gestures and their application in enhancing sign language recognition and speech synthesis. The system uses real time video capture for improved hand detection network which identifies positions of hands. Through modern algorithms feature extraction and representation are improved leading to more precise recognition of gestures. For example this information could be integrated with sign language systems and text-to-speech engines in order to make them more efficient improving natural human computer interaction possibilities. The system offers future possibilities for communication accessibility advancement and human



ai .

Figure 2. Sign language using ResNet

Initially captures the speech waveform, that's the analog illustration of sound waves. The next step is to transform the waveform into a digital signal the use of an analog-to-virtual converter (ADC). The virtual sign is then processed with the aid of a function extractor, which extracts functions from the sign that are relevant to speech reputation. These features are then utilized by a decoder to understand the words that were spoken. The decoder uses a language model to assist it pick out the maximum in all likelihood sequence of phrases, given the extracted capabilities. Finally, the diagnosed phrases are output as text.

The precise steps concerned in every of those tiers can range depending at the specific speech reputation device. However, the primary process is the identical for all speech reputation systems.

Here are some of the challenges that speech popularity structures face:

Background noise: Speech recognition systems can be tough to apply in noisy environments, as the noise can intervene with the speech sign.

Speaker variability: People communicate with different accents, dialects, and talking styles. Speech popularity systems want with a view to take care of this variability. Vocabulary obstacles: Speech reputation structures may not be able to apprehend all phrases, particularly uncommon or technical words.

Despite these challenges, speech popularity systems have become an increasing number of correct and dependable in latest years. They are utilized in a extensive range of programs, inclusive of dictation software, voice instructions for devices, and automatic captioning.

C. *SPEECH AND TEXT GENERATION USING TSS ENGINE*

The manner of converting spoken language into text involves several intricate steps, every contributing to the seamless transformation from the analog area of sound waves to a comprehensible virtual textual content output. Initially, the speech waveform, representing the analog model of sound waves, is captured. This raw illustration is then subjected to the analog-to-virtual conversion via an Analog-to-Digital Converter (ADC). This conversion is important for subsequent digital processing and analysis. The digitized signal undergoes in addition refinement within the shape of feature extraction. A feature extractor is hired to identify and extract applicable features from the digital signal that keep importance within the realm of speech popularity. These capabilities act as special markers, shooting the nuances and styles inside the spoken language. Following characteristic extraction, the processed statistics is surpassed over to a decoder, element chargeable for interpreting the spoken phrases. In the very last degree, the identified words are output as textual content. The culmination of correct waveform capture, analog-to-virtual conversion, function extraction, and language model-pushed interpreting outcomes in a dependable and coherent textual illustration of the in the beginning spoken phrases.

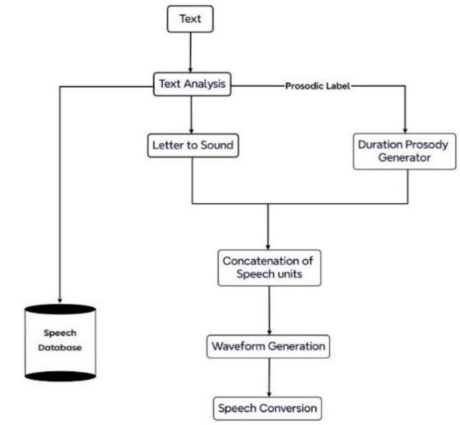


Figure 3. TTS engine

1.

II. PROPOSED ARCHITECTURE DESIGN

It outlines the person enjoy for interacting with our modern hand gesture translation system. Designed for accessibility and simplicity of use, the gadget permits you to translate spoken language via intuitive hand gestures, facilitating smoother conversation across languages.

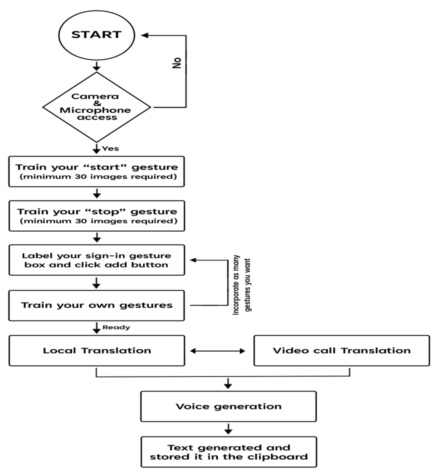


Figure 4. System architecture of sign language generation

1. Permission Granting:
Upon launching the software program, you may be precipitated to furnish get right of entry to to your camera and microphone. These permissions are essential for taking pictures hand gestures and audio appropriately during the interpretation manner.

2. Gesture Training:
To customize your enjoy, the tool calls for you to educate it three key gestures
Start: This gesture initiates a translation by way of sending the captured audio for processing.
Stop: Use this gesture to halt the modern-day translation and prepare for a new word.
Sentence End: This gesture indicates the crowning glory of a sentence inside the ongoing translation, improving interpretation and punctuation.
Training is easy and interactive. Follow the on-screen instructions to perform each gesture more than one times. The machine will learn and understand your hand actions for destiny accuracy.

3. Initiating a Translation:
Once your gestures are skilled, starting a translation is easy. Simply raise your hand and perform the Start gesture. The device will capture your spoken phrases and start the interpretation process.

4. Enhanced Translation Features:
During translation, you could utilize numerous treasured features activated by hand gestures:

Live Captions: Generate real-time text captions for both the authentic and translated speech, enhancing accessibility for members with hearing impairments.
English Translation: Convert spoken language into English, regardless of the authentic language, disposing of language obstacles.
Visual Descriptions: Receive text descriptions of applicable historical past scenes or actions, presenting context for visually impaired users.

5. Sentence End and New Gestures:
Break long speeches into manageable sentences and use the Sentence End gesture to complete a thought. This ensures accurate translation and punctuation. You can also customize the experience by adding new gestures beyond the initial set – see the user manual for detailed instructions on creating and training custom gestures

6. Ending a Translation:
To conclude a translation session, simply raise your hand and make the Stop gesture. The system will stop the current translation and prepare for a new one.

III. EXPERIMENT SETUP

A. RESULT AND EVALUATION

The performance and effectiveness of systems or models designed for translating and generating sign language content are calculated.
The face recognition is a technique to identify or verify the face from the digital images or video frame. A human can quickly identify the faces without much effort. It is an effortless task for us, but it is a difficult task for a computer. There are various complexities, such as low resolution, occlusion, illumination variations, etc. These factors highly affect the accuracy of the computer to recognize the face more effectively.

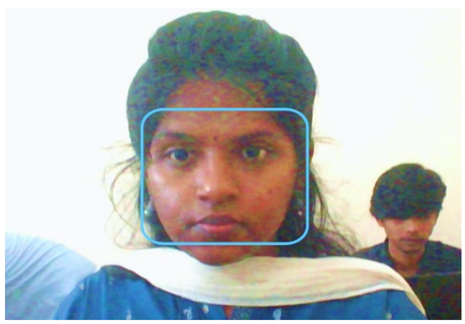


Figure 5. Face Detection and Face Localization

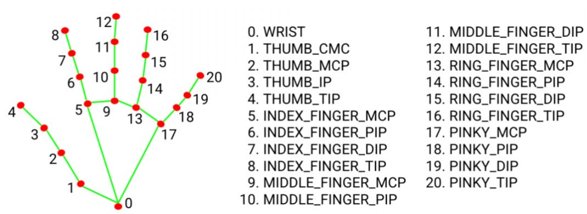


Figure 6. Hand landmark vectorization

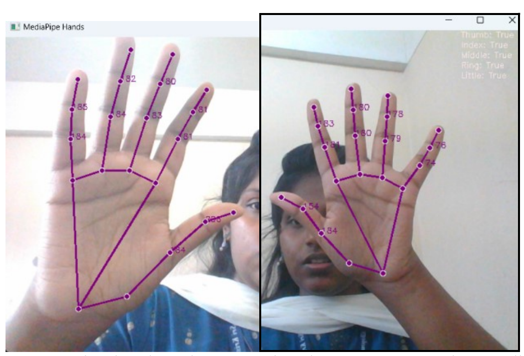


Figure 7. Detecting hands and representing them as vectors using Media Pipe

The hand land marker model bundle contains a palm detection model and a hand landmarks detection model. The Palm detection model locates hands within the input image, and the hand landmarks detection model identifies specific hand landmarks on the cropped hand image defined by the palm detection model.

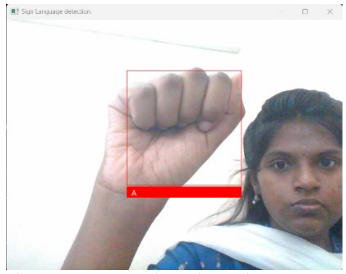


Figure 8. Gesture detection using ResNet8

Sign detection using ResNet8 involves training a powerful machine learning model to recognize hand gestures like used in sign language. ResNet8, known for its efficiency and accuracy, analyzes images or videos, identifying specific patterns linked to different signs. This approach can aid communication accessibility, translate sign language to text in real-time, and even power interactive interfaces, all fueled by a single model.

B. EVALUATION METRICS

The algorithm underwent a radical performance evaluation, utilizing the x-axis to signify don't forget (the proportion of real wonderful detections) and the y-axis for precision (the proportion of set of rules detections recognized as proper positives). A more superior set of rules is characterized with the aid of a higher curve on this context.

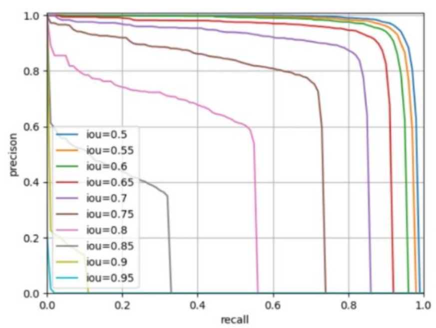


Figure 9. P-R curve of this algorithm under different IOU thresholds

In the left-hand graph, the blue curve illustrates the algorithm's overall performance throughout diverse IOU (Intersection over Union) thresholds, indicating the overlap among predicted and ground fact bounding packing containers. A better IOU threshold necessitates increased overlap for correctness.

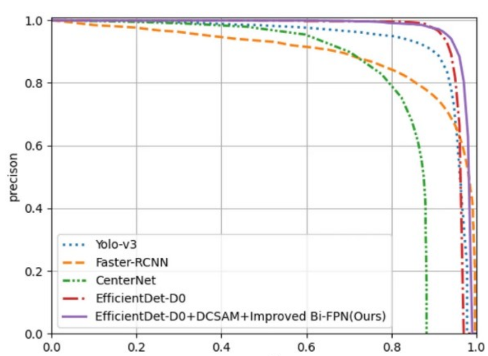


Figure 10. P-R curve of different detection algorithms when IOU = 0.5

The right-hand graph delves right into a contrast of the set of rules's overall performance at an IOU threshold of 0.5 with opportunity item detection algorithms, encompassing Faster R-convolutional neural community, CenterNet, EfficientDetDo, and EfficientDet D0+DCSAM+Improved Bi-FPN. Significantly, the blue curve in each graphs, representative of the set of rules in focus, surpasses the overall performance of its opposite numbers, mainly at higher IOU thresholds. This shows a great power of the algorithm in reaching particular detections in contrast to competing methodologies.

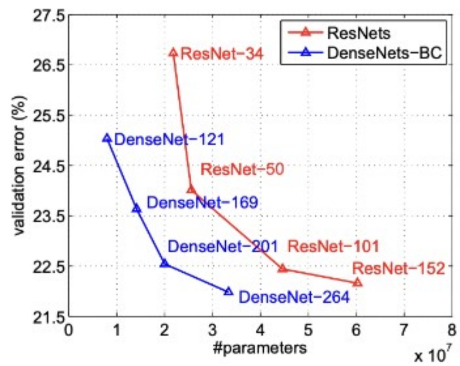


Figure 8. A Comparative Analysis of DenseNets and ResNets for Image Classification

In an analysis DenseNets have shown performance compared to ResNets consistently displaying lower validation errors while using the same number of parameters. This highlights the effectiveness of DenseNets in extracting insights, from a dataset. It's worth noting that the pattern of decreasing validation error as parameter numbers increase follows practices in learning indicating that these models can grasp complex data representations. However it's important to recognize the law of diminishing returns, where adding more parameters eventually leads to improvements in performance.

Specifically ResNet 152 stands out with the validation error among the architectures discussed although it comes with a computational load due, to its large number of parameters.

This underscores the trade-off between model accuracy and computational efficiency. The graphical representation strongly implies that DenseNets present a promising architectural choice for image classification tasks, showcasing notable accuracy levels while requiring fewer parameters than ResNets. The algorithm performed a radical performance evaluation, using the x-axis to indicate don't forget (proportion of actual surprising detections) and the y-axis for accuracy (proportion of set of rule detections recognized as appropriate positive) in this case with a higher curve to indicate more superior rule sets.

II. CONCLUSION

Our modern-day actual-time sign language era addresses important barriers within the present day scenario, presenting a promising revolution in verbal exchange for people with sensory disabilities. Utilizing a sophisticated version that incorporates ResNet and deep studying, our goal is to significantly improve sign language translation accuracy to tiers by no means seen before, accommodating various signing patterns seamlessly. The integration of sophisticated hand reputation capabilities and the incorporation of actual-time video call options through the TokBox API decorate the communique revel in, developing a unique and inclusive surroundings. These strides in innovation cross beyond mere technical enhancements; they represent a commitment to final current gaps, fostering connectivity, and improving accessibility for people who depend on sign language. Ultimately, our task pursuits to empower individuals in their daily interactions, making a transformative effect on the landscape of inclusive communication.

III. FUTURE SCOPE

The projects goal is to enhance the accuracy and efficiency of converting spoken words into text with a focus, on inclusivity and cutting edge technology. It involves exploring machine learning models, such as transformer based ones to improve speech recognition. Additionally there are plans to broaden language support tailor the system for languages and enhance real time adaptability using adaptive learning methods. Another aim is to boost word recognition accuracy by considering user intent and situational context for understanding. The project also aims to upgrade the user interface by integrating with emerging technologies like augmented and virtual reality utilizing edge computing for responses and collaborating with accessibility initiatives. These initiatives demonstrate the projects dedication, to addressing challenges and meeting the needs of users in an evolving technological landscape.

REFERENCES

[1] SIGNFORMER: Deep Vision Transformer for sign language recognition by deep r. Kothadiya 1, chintan m. Bhatt,tanzila saba 3, amjad rehman and saeed ali bahaj

[2] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, ''Deepsign: Sign language detection and recognition using deep learning,'' Electronics, vol. 11, no. 11, p. 1780, Jun. 2022.

[3] S. Katoch, V. Singh, and U. S. Tiwary, ''Indian sign language recognition system using SURF with SVM and CNN,'' Array, vol. 14, Jul. 2022, Art. No. 100141.

[4] A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu, and Q. Xin, ''Hypertuned deep convolutional neural network for sign language recognition,'' Comput. Intell. Neurosci., vol. 2022, pp. 1–10, Apr. 2022

[5] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, ''Skeleton aware multimodal sign language recognition,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2021, pp. 3413–3423.

[6] D. R. Kothadiya. Deepkothadiya/STATIC_ISL: Static Indian sign language dataset having sign of digit and alphabet. GitHub. Accessed: Oct. 19,2022. [Online].

[7] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[8] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[9] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[10] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[11] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[12] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[13] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[14] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[15] C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[16] M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[17] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[18] A. Thakur. (May 1, 2019). American Sign Language dataset. Kaggle. Accessed: Oct. 19, 2022. [Online]. Available: https://www.kaggle.com/ datasets/ayuraj/American-sign-language-dataset

[19] S. M. Rayeed. (Aug. 8, 2021). Bangla sign language dataset. Kaggle. Accessed:Oct.18,2022.[Online].Available:

[20] Nandini, Avni Verma, Sandeep Kumar, Sign and Voice Translation using Machine Learning and Computer Vision, JCSE, April 2023

[21] Obi, Yulius & Claudio, Kent & Budiman, Vetri & Achmad, Said & Kurniawan, Aditya. (2023). Sign language recognition system for communicating to people with disabilities.

[22] Saini B, Venkatesh D, Chaudhari N, et al. (2023) A comparative analysis of Indian sign language recognition using deep learning models

[23] Liang Z, Li H, Chai J. Sign Language Translation: A Survey of Approaches and Techniques. Electronics. 2023; 12(12):2678.

[24] Muhammad Aminur Rahaman, Kabiratun Ummi Oyshe, Prothoma Khan Chowdhury, Tanoy Debnath, Anichur Rahman, Md. Saikat Islam Khan,Computer vision-based six-

[25] layered ConvNeural network to recognize sign language for both numeral and alphabet signs, Biomimetic Intelligence and Robotics, Volume 4, Issue 1, 2024

[26] Bragg, Danielle & Verhoef, Tessa & Vogler, Christian & Morris, Meredith & Koller, Oscar & Bellard, Mary & Berke, Larwan & Boudreault, Patrick & Braffort, Annelies & Caselli, Naomi & Huenerfauth, Matt & Kacorri, Hernisa. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. 16-31. 10.1145/3308561.3353774. Bull, Hannah, Triantafyllos Afouras, Gül Varol,amuelnie,Liliane Momeni, and Andrew Zisserman. 2021. "Aligning Subtitles in Sign Language Videos." In Proceedings of the Ieee/Cvf International Conference on Computer Vision, 11552–61.