World Journal of Engineering Research and Technology
WJERT
www.wjert.org
SJIF Impact Factor: 5.924

SIGNSPEAK: BRIDGING COMMUNICATION THROUGH DEEP LEARNING

Sathvik Rao*, Shruthi D. V. (Assistant Professor)*, Sandeepa T. N., Anirudha R. and Keerthan V.

Department of Information Science & Engineering, Malnad College of Engineering Hassan.

Article Received on 09/02/2024 **Article Revised on 29/02/2024** **Article Accepted on 19/03/2024**

***Corresponding Author**
Sathvik Rao
Department of Information Science & Engineering, Malnad College of Engineering Hassan.

ABSTRACT

As the primary communication channel for individuals with hearing and speech impairments, sign language bridges the auditory gap through a nuanced tapestry of visual gestures and signs. However, seamless interaction between these individuals and the hearing population necessitates a shared understanding of the specific sign language dialect in use. Unfortunately, the widespread adoption of Indian Sign Language (ISL), characterized by its intricate blend of static and dynamic, uni- and bi-manual expressions, remains limited within the general populace. Further complicating this landscape are regional variations in ISL interpretations of even basic alphabetic symbols. This underscores the urgent need for technological interventions to bridge this persistent communication divide within the community. With this objective in mind, this study embarks on a comprehensive investigation of diverse approaches for ISL recognition. We delve into the intricacies of various image and video preprocessing techniques, including noise attenuation, segmentation algorithms, and feature extraction methodologies that capture the essence of these visual expressions. Furthermore, we explore the efficacy of established machine and deep learning algorithms in meticulously deciphering and accurately recognizing the dynamic vocabulary of ISL signs. This insightful survey illuminates the existing knowledge gaps and challenges that persist within the domain of ISL recognition, paving the way for future advancements in this critical field.

KEYWORDS: Hand gesture recognition, Convolutional neural network, Indian Sign

Language.

I. INTRODUCTION

Communication has perennially served as an intrinsic facet of human existence, where the capacity to articulate one's, thoughts remain foundational. Nevertheless, the realm of communication presents formidable hurdles for individuals contending with speech impediments,thereby necessitating their reliance on sign languageas a conduit for interaction. Sign language, a visual medium for information transmission, encompasses numerous linguistic variations globally, inclusive ofIndian Sign Language (ISL). ISL stands out for its inherent intricacy, owing to the amalgamation of both single- and double-handed gestural lexicons, aswell as the incorporation of both static and dynamicsignalling modes.

In India, a substantial demographic of approximately

19 lakhs grapples with speech impairments, underlining an imperative need for technology capable of deftly and precisely recognizing ISL signs and seamlessly rendering them into a human- comprehensible format. Eminent researchers have undertaken multifaceted approaches encompassing the domains of image and video processing, machinelearning, deep learning, and sensor-driven hardware mechanisms, with the overarching objective of engendering robust ISL recognition systems.

The core focus of this scholarly endeavour resides in the succinct encapsulation of technological innovations germane to ISL recognition, while simultaneously accentuating lacunae and challenges entrenched in the current corpus of knowledge. This comprehensive survey aspires to proffer valuable insights, which, in turn, will serve as a compass for those navigating the landscape of knowledge dissemination and solution implementation in addressing conundrums and contingencies with innovative predispositions.

II. LITERATURE SURVEY

[1]A prior study unveiled "Mudra," a pioneering Indian Sign Language (ISL) recognition system specializing in deciphering dynamic gestures pertinent to the banking domain. Mudra operates on a bespoke database meticulously curated to encompass 20 banking-specific signs and a selection of commonly used signs. This meticulously captured dataset comprises 1100 video recordings of varying lengths, generously contributed by student volunteers, with each video boasting a rapid frame rate of 40 frames per second. The dataset is judiciously divided into training and testing sets, adhering to a well-established 80:20 ratio.

For feature extraction, Mudra leverages the formidable InceptionV3 convolutional neural network (CNN) architecture. This architecture incorporates a rectified linear unit (ReLU) correctionlayer, a max-pooling layer, and two fully connected layers, meticulously extracting intricate visual attributes from the input signs. The output generated by this CNN model is then meticulously fed into a Long Short-Term Memory (LSTM) network, responsible for the critical task of symbol classification and subsequent conversion into textual representation. Importantly, LSTM's inherent ability to obviate the need for manual feature engineering positions it as the preferred choice compared toalternative deep learning methodologies. The meticulously designed architecture demonstrably achieves a remarkable 100% training accuracy, coupled with a commendable 81% testing accuracy, showcasing its efficacy in deciphering the financial lexicon of ISL.

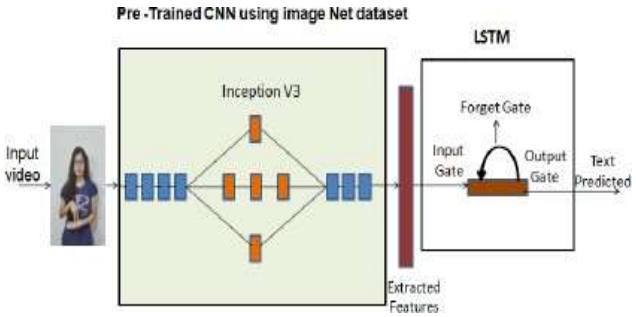

Figure 1: Training CNN model.

The paper[2] delves into the innovative utilization of MediaPipe technology for real-time recognition of hand gestures, with a particular emphasis on its applicability to sign language, adomain where it holds immense promise for individuals grappling with hearing impairments. The authors detail their utilization of MediaPipe's robust library, enabling the precise prediction of a human hand's skeletal structure and intricate gestures. This precision is realized through the integration of two key models: a palm detector and a hand landmark model.

A noteworthy focal point of this study is the quest for achieving lightweight and resource-efficient hand gesture detection, ideally suited for mobile devices. The authors employ rigorous quantitative analysis, punctuated by comparisons to alternative methods, notably the deployment of Support Vector Machine (SVM) algorithms. The outcome of these comparisons showcases the exceptional prowess of their model, manifesting in an average accuracy rate of a staggering 99% across multiple sign language datasets. In addition, the study accentuates the cost- effectiveness, real-time responsiveness, and adaptability of their

model when confronted with diverse sign language datasets.

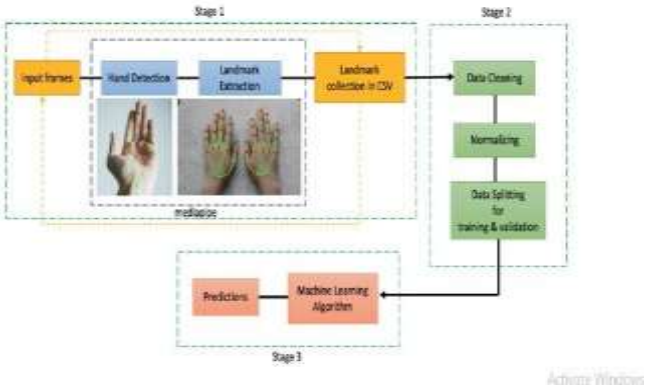

Figure 2: Gesture Detection.

[3]A recent foray into the realm of Indian Sign Language (ISL) recognition unveiled "DeepSign," an advanced deep learning paradigm that leverages the prowess of Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) architectures. This groundbreaking system meticulously decipherssign language gestures embedded within video streams, seamlessly translating them into their corresponding English counterparts.

The meticulously curated IISL2020 dataset, acollaborative effort by 16 participants (both male andfemale, aged 20-25), serves as the foundation for DeepSign's learning. Encompassing 11 distinct words, the dataset boasts a total of 1100 video samples per word, each artfully captured using mobile devices. These high-resolution (1920 x 1080)video samples, recorded under natural lighting conditions with an impressive 28 frames per second, accurately capture the nuances of hand movement and expression. Each video segment, averaging 2 seconds in duration, provides a glimpse into thedynamic nature of ISL communication.

The computational muscle of a 16GB GDDR6 GPU powers the model training process. Feature extraction is adeptly conducted through meticulous sampling methodologies, extracting the salient characteristics from individual video frames. These extracted features are subsequently processed through a pre-trained MobileNet coupled with Inception ResNetV2, culminating in the generation of a feature vector that feeds into the LSTM-based predictions. To ensure the model's robustness and generalizability, a rigorous ten-fold K-fold cross-validation methodology is employed.

A recent delve into Indian Sign Language (ISL) recognition unveiled two distinct methodologies proposed by astute researchers.[4] The first approach embarks on a meticulous

segregation of hand movements, meticulously extracting these intricate nuances from both depth and RGB data. This initial method focuses on a curated repertoire of 36 static signs, akin to individual frames captured ina silent film. Affine transformation and 3D construction techniques are adeptly wielded to parse the data, meticulously separating the hand gestures from the background scene. These segmented hand regions are then fed into powerful Convolutional Neural Networks (CNNs), akin to digital brains trained on visual patterns. The culmination of this intricate process is an impressive classification accuracy of 98.91%, showcasing the efficacy of this approach in deciphering the static vocabulary of ISL. The second methodology takes a different path, focusing on the dynamic nature of sign language communication. It harnesses the prowess of Long Short-Term Memory (LSTM) networks, renowned for their ability to learn and process temporal sequences. Convolutional kernels further augment this learning process, enabling the system to extract salient features from the video data. This approach tackles a dataset encompassing 10 dynamic signs, akin to a short video clip showcasing the flow of movement within a sign. It culminates in a remarkable classification accuracy of 99.08%, demonstrating its efficacy in capturing the essence of dynamic ISL expressions. Notably, this methodology incorporates the innovative U-Net Architecture, which eliminates the need for a specialized RGB-D Kinect camera. This opens doors to increased accessibility and broader potential applications, as it frees the system from dependence on specific hardware requirements.

[5]A recent foray into the realm of Indian Sign Language (ISL) recognition witnessed the unveiling of a groundbreaking Convolutional Neural Network (CNN)-based ISL converter.[5] This cutting-edge architectural marvel boasts remarkable proficiency in deciphering the 26 alphabetic symbols that form the bedrock of ISL communication. The authors leverage the power of transfer learning, a paradigm where knowledge gleaned from pre-trained models is strategically harnessed to accelerate learning in new domains. In this instance, they meticulously retrain the final layer of the pre-trained MobileNet model, transforming it into a potent classifier for ISL alphabets.

The foundation of this system lies in a self-curated dataset, a veritable treasure trove encompassing astaggering 52,000 images. Each alphabet symbol ismeticulously represented by 2000 meticulously captured images, ensuring comprehensive coverage and diversity. To further enrich the dataset'srepresentational power, the authors employ a repertoire of augmentation techniques. These techniques, akin to digital artists manipulating their

creations, introduce variations such as background alterations, image cropping, flips, expansions, andresizing. This meticulous process injects valuablediversity into the dataset, mimicking the real-worldvariations encountered in sign language expression. To accurately segment the hands from thebackground, the system seamlessly integrates the prowess of the GrabCut Algorithm. This algorithmic maestro adroitly isolates the crucial hand gestures, akin to a sculptor removing excess marble to reveal the hidden masterpiece within. The culmination of this intricate process is an impressive testingaccuracy of 96%, showcasing the system's remarkable ability to decipher the alphabet of ISL expression.

This seminal research paper[6] delves into the intricate workings of a real-time interactive system that bridges the gap between gesture recognition and phrase generation within the dynamic sphere of Indian Sign Language (ISL). Functioning akin to a skilled interpreter, this system seamlessly processes video data laden with ISL gestures and transforms them into coherent, grammatically sound phrases.

Fueling this intricate system is a wellspring of knowledge in the form of a vast dataset encompassing 10,000 carefully curated images. Each image serves as a window into the nuanced vocabulary of ISL, meticulously capturing 100 unique signs in four distinct formats: raw, FAST, Canny Edge, and SIFT. To prepare this dataset for the analysis, a meticulous array of sophisticated preprocessing techniques are meticulously applied, akin to a chef diligently preparing ingredients for a masterpiece. The heart of the system lies in a meticulously trained hybrid Convolutional Neural Network (CNN) model, a veritable maestro adept at deciphering and classifying the intricate tapestry of ISL gestures. Once a gesture is accurately identified, it serves as the key unlocking a treasure trove of meaningful phrases. Employing this key with finesse, the system ingeniously crafts grammaticallysound phrases, thoughtfully weaving them with relevant words to illuminate the underlying communication intent.

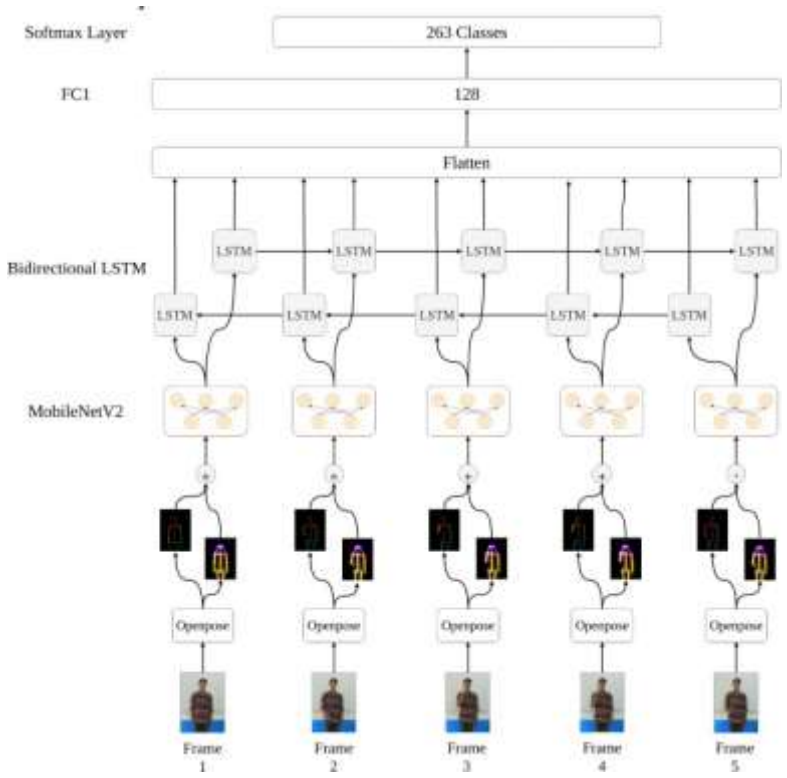

Figure 3: CNN weights.

In a seminal contribution to the field of Indian Sign Language (ISL) recognition,[7] researchers unveiled a groundbreaking advancement: the INCLUDE dataset. This extensive, open-source treasure trove boasts a staggering 263 distinct words, serving as a valuable resource for researchers and developers alike.

The system's robust methodology for recognizing multiple sign languages seamlessly integrates data augmentation and feature extraction during the preprocessing stage. The INCLUDE dataset undergoes meticulous curation, meticulously diversifying its representations through techniques such as horizontal and vertical image flipping, cropping, and size alterations. Feature extraction is skillfully conducted utilizing pre-trained models like OpenPose, Pose Videos, and PAF Videos. These extracted features are then meticulously flattenedand normalized before being fed into an array of machine learning classifiers.

The INCLUDE dataset truly shines when comparing different algorithms. Remarkably, the XG-Boost algorithm outperforms common recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, demonstrating its superior efficacy in this domain.

For gesture recognition, the system leverages Pose and PAF videos, extracting features with the aid of the pre-trained MobileNetV2 model. These extracted features are then meticulously

channeled through a BiLSTM architecture for classification. The hidden states of the LSTM cells are flattened and conveyed through a fully connected layer and softmax layer, culminating in an impressive overall accuracy rate of 85.6% on the extensive INCLUDE dataset.

In this deep learning,[8] the focal point was the utilization of the CIFAR-10 dataset as the cornerstone for image classification. A comprehensive preprocessing pipeline was meticulously executed, encompassing the pivotal stages of data augmentation for enriched dataset diversity and normalization for optimized data scaling and distribution. The core architectural paradigm adopted was a Convolutional Neural Network (CNN), strategically configured with two convolutional layers meticulously engineered for feature extraction and abstraction. This was supplemented by a singularly sophisticated fully connected layer. The culmination of this technical pursuit was characterized by the assessment of recognition accuracy on the exacting test set, yielding a commendable performance benchmark, which consistently approximated an 85% threshold.This outcome underscores the model's technical acumen in effectively categorizing a diverse array of images, thereby solidifying its pre-eminence in the specialized domain of image classification.

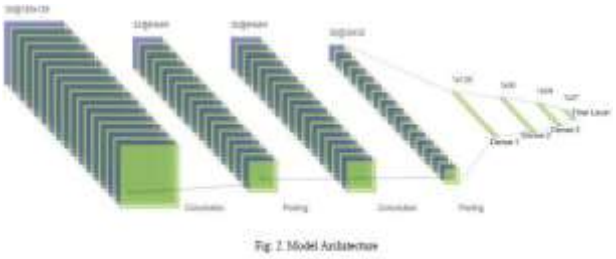

Figure 4: CNN layers.

III. RESEARCH GAPS

Our comprehensive survey of existing research illuminates critical lacunae within the domain ofIndian Sign Language (ISL) recognition. Although widely used, ISL suffers from a dearth of dedicated research compared to prominent sign languages like American Sign Language (ASL). This disparity is further exacerbated by the inherent complexity of ISL, characterized by the intricate interplay of manual and non-manual features in its gestures. Notably, ASL primarily relies on single-handed expressions, posing a distinct challenge for automated recognition systems.

Compounding this issue is the conspicuous absence of a standardized dataset specifically

tailored for capturing the dynamic nature of ISL signs. This deficiency forces researchers to create their own datasets, often limited in size and scope. Additionally, several existing ISL datasets are plagued by inconsistencies, including poorly defined gesture lengths and suboptimal image quality due to inadequate lighting conditions. Furthermore, the distinct vocabulary and grammatical structures of ISL render techniques developed for other international sign languages largely inapplicable.

Our research endeavors are strategically designed to address these prevalent gaps, paving the way for the development of robust and effective ISL recognition systems. Our efforts aim to contribute significantly to bridging the communication divide and empowering individuals who rely on ISL for expression and interaction.

IV. CHALLENGES

The quest for efficacious and precise Indian Sign Language (ISL) recognition systems demands innovative solutions capable of navigating a labyrinth of multifaceted challenges. Paramount among these hurdles lies the inherent complexity of gesture patterns within video datasets. The intricate choreography of both hands, the delicate interplay between hand and body, and the subtle dance with environmental nuances pose significant hurdles, further hampered by potential finger and hand occlusions. This intricate tapestry of complexities renders traditional recognition methods ill-equipped, demanding sophisticated approaches to unveil the true meaning buried within these dynamic expressions.

Furthermore, ISL's compounding sign system, where the meaning transcends individual gestures, multiplies the intricacies involved in deciphering its nuanced language. The delicate interplay between background context and camera technology throws another curveball, necessitating meticulous consideration of factors like resolution and orientation during system design.

Adding to the complexity, the dream of continuous Sign Language Recognition (SLR) demands a modeling framework that not only captures the fleeting gestures but also grasps the idiosyncrasies inherent to the language itself. Managing long-term sequential data with its inherent computational burden, coupled with the labyrinthine linguistic rules of ISL, elevates the challenge to unprecedented heights.

Yet, despite the formidable obstacles, the imperative for robust and precise ISL recognition

systems remains undimmed. Through the adroit navigation of these challenges and the pioneering of innovative techniques and methodologies, the realization of a system that faithfully translates the dynamic symphony of ISL gestures into its intended meaning becomes a tangible reality. Such advancements hold the transformative power to bridge communication gaps and empower individuals who rely on sign language, weaving a more inclusive tapestry of understanding for all.

V. OUTCOMES

Table i: Some commonly used techniques for dataset creation, preprocessing, and architectures along with the values of evaluation parameters.

Reference	Dataset	Preprocessing Techniques	Architecture	Recognition Accuracy
[1]	Self-created dataset of20 bank-related and everyday words. 1100 videos. Recorded on Mobile at 40 fps	CNN (InceptionV3) for feature extraction. Layers used - conv2D, Maxpooling2D, Avg Pooling.	LSTM (Long Short-Term Memory)	Training - 100% Testing - 81%
[2]	Multiple sign language datasets (ASL, Indian Sign Language, ItalianSign Language)	MediaPipe library for hand landmarks - Data cleaning, normalization, and splitting	Two-stage pipeline Support Vector Machine (SVM) for recognition	SVM achieved an average accuracy of 99% for most sign language datasets
[3]	Self-Created dataset of11 signs. 1100 videos per sign. 16 subjects contributed to creating the database which was recorded at 20 fps on a mobile phone.	Video frames areresized. InceptionResNetV2 isused for feature extraction from these frames.	GRU-GRU LSTM-LSTM GRU-LSTM LSTM-GRU	Training- 97%
[4]	5 subjects of both genders volunteered fordataset creation. The videos were captured using 9 different rates. Approach 1: 90 thousand RGB-D images were captured using a Microsoft Kinect camerafor 36 static signs + augmentation. Approach 2: 1080 videos of 10 dynamic signs.	Approach 1: 3D Reconstruction + affine Transform Approach 2:U-Net semantic segmentation	Approach 1: Basic CNN Approach 2: LSTM	Approach 1: Training accuracy - 98.81% Approach 2: Training accuracy - 99.08%
[5]	Self-created dataset of 10000 images for 100signs that include	Gray scaling,illumination normalization, noise removal, edge detection,	Hybrid-CNN	-94.2%

	alphabets A-Z, digits 0-9, and 64 frequently used words	corner detection, thresholding		
[6]	Self-created Dataset (INCLUDE). Publicly available 263 signs from15 different word categories. The deaf community has helped in recording the dataset. (7 senior students). A total of 4287 videos. Signs are of 2-4 second length. The video resolution is 1920x1080 and the frame rate is 25 fps.	Center crop, Horizontal flip, Up-sample, and Down-sample for preprocessing. MobileNetV2 was usedfor feature extraction.	XGBoost BiLSTMs	Accuracy - 94.5% (50 words) Accuracy - 85.6% (263 words)
[7]	Self-Created dataset of 12 people including an expert. The images werecaptured at 480x640 resolution with normal lighting. A total of 312sign images of 26 alphabets were captured.	Image resizing, Hang segmentation. HOG and canny edge detection for feature extraction.	Extreme Machine Learning (EML)	Testing accuracy - 80.76%
[8]	CIFAR-10 (Image Classification)	Data Augmentation, Normalization	CNN with 2 Conv Layers, 1 FC Layer	Test Accuracy: ~85%

CONCLUSION

In summation, our comprehensive examinationencompassing over 50 scholarly publications concerning Indian Sign Language (ISL) recognition underscores the imperative to confront the distinctive challenges intrinsic to this linguistic domain. These challenges primarily revolve around its inherently multimodal character, intricate anddiverse gesture patterns, and linguistic idiosyncrasies. Within our scrutiny, a plethora of preprocessing methodologies have come to the fore, ranging from Hand Segmentation to Mask Generation. Additionally, we have discerned the pivotal role played by popular deep learning and machine learning paradigms, including Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Naive Bayes, all of which have demonstrated prowess in ISL recognition.

While commendable progress has been realized, the journey towards enhanced recognition accuracy and heightened accessibility for sign language users necessitates a continued commitment to further research and innovation. By pioneering bespoke algorithms and methodologies tailored to the intricate nature of ISL, the prospect of crafting precise and effective recognition systems for IndianSign Language comes into sharper focus.

REFERENCES

1. Jayadeep, Gautham, et al. "Mudra: convolutional neural network based Indian sign language translatorfor banks." 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2020.
2. Kavana KM, Suma NR"Recognization Of Hand Gestures Using Mediapipe Hands" IRJETS, 2022; 04: 2582-5208.
3. Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K.,Gil-González, A. B., & Corchado, J. M. Deepsign: Sign language detection and recognition using deep learning. Electronics, 2022; 11(11): 1780.
4. Likhar, Pratik, Neel Kamal Bhagat, and G. N. Rathna. "Deep learning methods for Indian sign language recognition." IEEE 10th InternationalConference on Consumer Electronics (ICCE- Berlin). IEEE, 2020.
5. Gangadia, D., Chamaria, V., Doshi, V., & Gandhi, J. (December). Indian sign language interpretation and sentence formation. In IEEE Pune section international conference (PuneCon), 2020; 71-76. IEEE.
6. Sridhar, A., Ganesan, R. G., Kumar, P., & Khapra, M. (October). Include: A large scale dataset for indian sign language recognition. In Proceedings of the 28th ACM international conference on multimedia, 2020; 1366-1375.
7. Kumar, Anand, and Ravinder Kumar. "A novel approach for ISL alphabet recognition using Extreme Learning Machine." International Journal of Information Technology, 2021; 13: 349-357.
8. Yulius Obia, Kent Samuel Claudioa, Vetri Marvel Budimana, Said Achmada, Aditya Kurniawana. "Sign language recognition system for communicating to people with disabilities." ProcediaComputer Science, 2023; 216: 13–20.