Journal of Advanced Artificial Intelligence
Volume 1 – No.7, May 2025

Development of Computational Models for Automatic Indian Sign Language (ISL) Learning and Gesture Recognition

Mehuli Mukherjee
Department of Computer Application,
PES University, Bengaluru, India
ORCID: https://orcid.org/0000-0001-6964-3598

ABSTRACT
Indian Sign Language (ISL) is a crucial task in the areas of computer vision and pattern recognition, having its own grammar, syntax, vocabulary, and several unique linguistic attributes. It has wide applications in different aspects but the environment, background image resolution, modalities, and datasets affect the performance a lot. Over 466 million people are speech or hearing impaired, and 80% of them are semiilliterate or illiterate according to the WHO (World Health Organization) reports. So far discrete sign language videos are used to train the models. An enhancement opportunity has been proposed, where publicly available continuous videos can be considered and then using masked model algorithms to achieve a continuous conversation of speech/audio. Several deep neural networks methodologies are used which combines different methods for hand movement tracking, feature extraction, encoding and decoding and machine learning modeling. The current state of art is also leveraged for translating into different Indian languages. Finally, a comparative study is proposed to compare between widely used American Sign Languages.

Keywords
Indian Sign Language, Deep Neural Networks, Machine Learning

1. INTRODUCTION
According to the statistics of WHO (World Health Organization), approximately 466 million people are speech or hearing impaired, and 80% of them are either semiilliterate or illiterate. Around 5% of the world population suffers from a lack of hearing power. As per the prediction of the United Nations, the number of deaf people in 2050 will be 900 million [1]. Non-verbal manner conveys and communicates our views, emotions, and thoughts visually through sign language. Sign language grammar is quite different when compared to spoken language. A sign is associated with specific movement of hands, shapes, or signals produced in a particular location on or around the signer's body. Sign language recognition (SLR) is a complex and challenging domain where many research opportunities are available with the present technology of artificial intelligence and machine learning. It comprises various datasets, input modality, features, classification algorithms, computational resources, and other applications. The dataset is further classified into mainly two broad categories - isolated sign dataset and continuous sign dataset. Input modality comprises vision-based modality and sensor-based modality. The major features that concern SLR are hand movement, facial expression, and body movement.

Classification is categorized into traditional methods like (HMM, RNN, etc.), deep learning (CNN), and hybrid method which is a combination of traditional and deep learning or combination of optimization and deep learning algorithm.

There is no such comprehensive review paper addressing the aspect of the modality (vision and sensor), different types (isolated (manual and no manual) and continuous (manual and no manual)), various sign language datasets, and stateof-the-art methods-based studies. The SLR involves the process of data collection, preprocessing, feature extraction, and classification phase. These stages are discussed in detail below.

Vision and Sensor are the two methodologies to perform the data acquisition in SLR. The input is an image or video [2], [3] in a vision-based approach. To collect standard signs a single camera is used while active and invasive devices help to collect the depth information of multiple cameras. To capture the continuous motion, video camera, webcam, or smartphone device [4], [5], [6], [7] are used. The sensor is used to collect the signal [8], [9], [10], [11] in a sensor-based approach.

The performance and accuracy of the SLR system can be improved by the preprocessing methodologies such as dimensionality reduction, normalization, and noise removal procedures [12].

The image is splitted into various parts or ROI (Region of Interest) [13], Skin Color Segmentation [14], HTS (Hands Tracking and Segmentation) [15], Entropy Analysis and PIM (Picture Information Measure) [16] in the segmentation stage. The background also requires the hand gesture extraction to be done effectively with the help of segmentation and tracking process.

Tracking of facial expression and hand position from the acquired video/image can be performed using different algorithms and methodologies like camshaft (continuously adaptive mean shift used to track the head position) [17],

Adaboost with HOG (Histogram of the gradient) [18], Particle filtering (KPF-Kalman Particle Filter) [19].

The method of transforming preprocessed input data into the feature space is known as feature extraction. Feature extraction plays a vital role in SLR. It is vital and crucial for sign language detection, as irrelevant features may lead to misclassification [20]. The feature extraction also aids in speed [21] and improvement of accuracy. SURF (Speeded Up Robust Feature) [22], speed up robust feature (Laplace of Gaussian with box filter) [23], SIFT (shift-invariant feature transform) [24], PCA (Principal Component Analysis) [25], [4], LDA (Linear Discriminant Analysis) [26], Convexity defects and k-curvature [27], time domain to frequency domain [28], [29], Local binary pattern, etc.

are some of the feature extraction methods. PCA, LDA, etc. dimension reduction processes reduce the computational burden on the classifiers. The features reduction, dimensionality pruning, and lowering of the dimension helps to keep the significant features of high variance and minimizing remaining features hence reducing the training complexity.

The preprocessing and feature extraction methods also reduce the computation burden. Overfitting issues are taken care of, and wrong recognition possibilities are eradicated.

SIFT's merits are invariant to lighting, orientation, and scale though the performance is not satisfactory [30]. The unwanted information is removed by keeping the significant features to ease the image processing by using Histogram of Oriented Gradients (HOG) [31]. Using the computation of gradient margin and angle the feature vectors are obtained. The extracted feature also increases due to the increase of HOG cell size and the number of bins. Small subdivisions given local information that is worthwhile, larger subdivisions furnish the global information. The critical demerits of both the methods are that they require more memory for the computational process. SURF is invariant to image transformation and even a faster feature extractor compared to SIFT. The requirement of camera setup in horizontal position for better performance, and the disadvantage is illumination-dependent, not rational. Discrete Wavelet Transform is used to capture Location and frequency. Temporal resolution is considered to be the critical merit of DWT [32], [33].

The collected data (image/video) is stored in the database and classified into two broad sets, namely training and testing datasets [34]. The classifier learns by training a dataset and the performance is evaluated based on the testing data.

The classifiers perform the classification step by extracting features and hence classifying the sign gesture. Classifiers like the Hidden Markov Model (HMM) [9], Long-Short Term Memory (LSTM) [22] Deep Learning network [20], and hybrid classifier [2], [24] are used to recognize sign language. Sign dependent and sign independent SLR using three datasets using single and fusion parallel 3DCNN were performed by Al-Hammadi et al. [20]. CNN and LSTM based SLR models for Turkish SLR are performed by Sincan and Keles [32]. The feature extraction process is improved by FPM (Feature Pooling Module), convergence speeds up using the attention model which has a vital role in SLR. DCNN (deep convolution neural network and LSTM (long short-term memory) based model for hand gesture recognition has been used by Yuan et al. [33]. The gradient vanishing and overfitting problem is overcome by the residual module. The complicated hand gesture longdistance dependency problem has been addressed by an improved deep feature fusion network. Compared to algorithms like Bayes, KNN, SVM, CNN, LSTM, and CNNLSTM, the DFFN based model performs better on ASL and CSL datasets. An Arabic SLR model using BiLSTM (deep Bi-directional Long Short Term Memory recurrent neural network) has been designed by Aly and Aly [2]. The suggested model proved validity on signer independent real Arabic SLR. A multi-modal and multi-view hand skeletonbased SLR model is carried out by Rastgoo et al. [3]. Features fusion and single-view vs. a multiviewprojection of hand skeleton-based performance analysis was carried out. To recognize the hand sign language automatically, SSD (Single Shot Detector), 2DCNN (2D Convolutional Neural Network), 3DCNN (3D Convolutional Neural Network), and LSTM (long short-term memory) based deep pipe-line architectures were proposed. The kNearest-Neighbour method associated with Long-Short Term Memory (LSTM) recurrent neural network-based American SLR model was designed by Lee et al. [22] . The leap motion controller is utilized to achieve the sign data. The proposed model (LSTM with KNN) outperforms 99.44% when compared to SVM, RNN, and LSTM models.

2. DATASETS ON SIGN LANGUAGE

Across the world, several datasets have been proposed for sign languages out of which few are related to Indian Sign Language. There are not too many data sources available for Indian Sign Languages. In this state of art the INCLUDE - Indian Lexicon Sign Language Dataset [34] - for ISL is considered. It is a publicly available ISL dataset and compares favorably with datasets on other sign languages as well. Each class corresponds to a single sign in ISL. These classes belong to 15 major categories, covering popular sets of daily usage words in ISL. Some of the categories are in Table 1: Adjectives, Colors, Days and Time, Electronics, Animals, Clothes, Objects at Home, Occupations, People, Places, Pronouns, Greetings, Means of Transport, Seasons, Society. Majorly the signs are 2-4 seconds in length, having an average duration of 2.57 seconds. The resolution of each video is 1920x1080, with a frame rate of 25 fps.

The INCLUDE dataset has 0.27 million frames covering 263 different signs in total.

It can be computationally expensive to train deep learning models on a large video dataset. For faster evaluation of different models, a smaller subset called INCLUDE-50 with 50 of the 263 classes are considered. The classes chosen are: Bank, Bird, Black, Court, Cow, Death, Dog, Election, Boy, Brother, Car, Paint, Pen, Priest, Red, Shoes, Shop, Summer, T-Shirt, Teacher, Cell phone, Fall, Fan, Father, Girl, Good Morning, Hat, Hello, House, I, Monday, Thank you, Time, White, Window, Year, Large, Dry, Good, Happy, Hot, It, Long, Loud, Short, Small, Train Ticket, You (plural). New, Quiet. The classes cover all 15 categories and have been selected on the basis of frequency of usage. As a whole, INCLUDE-50 contains 958 videos and 60897 frames, and is approximately onefifth of the main INCLUDE dataset.

The data base for this study has been extended by using the huge repositories of publicly available data of ISLRTC [35] which is the Department of Empowerment of Persons with Disabilities, under the Ministry of Social Justice and Empowerment Under the Government of India.

ISLRTC launched the first set of Indian Sign Language Dictionary of 3000 terms on 23rd March, 2018 at India International Center, New Delhi. The Dictionary was released in DVD form, containing signs of everyday use and their corresponding English and Hindi words. Specialized terms from legal, academic, medical, and technical fields are explained in ISL. Indian Sign Language Research and Training Centre (ISLRTC) launched the 2nd Edition of Indian Sign Language Dictionary on 27th February, 2019.The second edition includes 6000 words under the categories of academic, legal, medical, technical and everyday terms.Indian Sign Language Research and Training Centre (ISLRTC) inaugurated the 3rd Edition of Indian Sign Language Dictionary with 10,000 terms on 17th February 2021. There are approximately 500 Technical Terms, 400+ Legal Terms ,1300+ Everyday Terms ,800+ Agricultural Terms and over 1000 Academic Terms. A glimpse of these terminologies is shown in Table 1.

Data is also collected from Indiansignlanguage.org[36]

(Faculty of Disability Management and Special Education

(FDMSE) of Ramakrishna Mission Vivekananda University (RKMVU), Coimbatore in collaboration with C-DAC

Hyderabad), which offers a huge collection of Indian Sign Language (ISL) signs. It's a non-profit organization in the field of special education and rehabilitation of persons with disabilities in India with more than 50 years of dedicated service. Each sign has an image, running video and threaded discussions. It consists of around 70 various categories like Banking, Art and Entertainment, Beverages, Cities and Towns etc. Some of them are shown in Table 1.

Table 1: Sample categories of terminologies from the three different data sources

Data Source	Categories
INCLUDE	Animals
	Cloths
	Colors
	Days and Time
	Occupations
	People
	Places
	Seasons
	Society
	Pronouns
ISLRTC	Technical
	Legal
	Agricultural
	Academic
	Everyday Terms
RKMVU	Banking
	Art and Entertainment
	Beverages
	Cities
	Towns
	Body Parts and Functions
	Dishes and Spices
	Heat

The process of converting pre-processed input data into a feature space is referred to as feature extraction. This step is crucial in Sign Language Recognition (SLR) and has a significant impact on its accuracy.

3. METHODOLOGY

Figure 1. OPENPOSE Tracking Data

In this section the proposed framework for SLR on the out combined data is discussed. The data collected is in the form of video files from data sources and for uniformity some preprocessing steps have been used followed by a feature extraction steps before feeding into the machine learning pipeline.

- Pre processing
- Feature Extraction
- Machine Learning Pipeline

The overall methodology has been summarised in Figure 2.

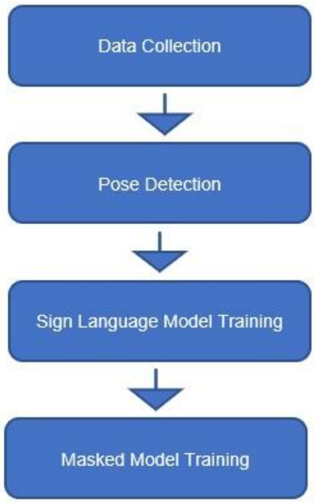

Figure 2. Overall methodology for Sign Language Detection

A. Pre-Processing

The current work tries to combine different dataset already available [] to achieve the goal of sign language recognition. In order to this there are two aspects that has been considered:

Uniformity

Three different data sources have been used with each with different resolution, frame rates and frame size. A uniform sampling of 10fps has been adapted for sampling all of the video's frame and minimum frame size has been set to accommodate all the videos from different sources.

Augmentation

To increase the number of data points for the model training process the standard approach that is adapted is oversampling the dataset. This normally means creating variation of the data while preserving the mean(statistical). In the current scheme of work cropping, sampling, flipping, and sampling have been used to oversample the dataset after the feature extraction step.

B. Feature Extraction

For developing deep learning models for motion prediction, a requirement of motion estimation framework has been long felt to extract key features in vector form from a video stream. Significant effort has been invested towards the direction of real time multi-person 2 D pose estimation. OpenPose [X] is one of the recent works that targets to achieve this objective by employing bottom-up non-parametric representation, named as Part Affinity compared to totopdown approach. The pre-trained model actually returns around 135 key-points in any image to identify different body parts present in a video. Open Pose Models returns three sets of feature from an incoming video stream namely:

- PAF Video
- Pose Video
- Key-points vector

All the three have been explored in the current scope of work. C.Machine Learning Pipeline

The proposed machine learning framework combines two different machine learning frameworks to combine Continuous and Discrete gin language into a single framework.

XGBOOST Classifier

The features generated from the OPENPOSE model is fed to the XGBOOST [38] classifier for a multi-class classification problem.

Tree boosting mechanism is a very effective and widely used machine learning method. XGBoost is a scalable end-to-end tree boosting system widely used to achieve state-of-the-art results on many machine learning problems. It is a sparsity aware algorithm for sparse data and weighted quantile for approximate tree learning. XGBoost can be executed very fast on standard CPUs. The use of XGBoost in SLR supports the claim that XGBoost works well across data-types including temporal data.

BERT masked language model

The output from the classified is fed to a pre-trained masked model for sentence formation.

Language model has been effective for betterment of many natural language processing tasks. The two existing strategies for applying language model representations: feature-based and fine-tuning. They target to predict the relationships between different sentences by analysing them and taken into account the tokenization known as named entity recognition. This is also applicable to question answering, where models are required to produce fine-grained output at the token level. The fine-tuning-based approaches are improvised using BERT: Bidirectional Encoder Representations from Transformers [41]. BERT overcomes the unidirectionality constraint by using a “masked language model” (MLM) algorithm. The algorithm of the masked language model is such that it will randomly masks some of the tokens from the input, and then predict the original vocabulary of the masked word depending on its context. The MLM combines the left and the right context, which allows it to pretrain a deep bidirectional Transformer. Apart from the masked language model, “next sentence prediction” task is also used to jointly pertain to the text-pair representations.

4. RESULTS AND DISCUSSION

Classical machine learning models such as Support Vector Machines [39], XGBoost [38], Random Forests [40] ,LightGBM, Adaboosthaving additional depth information, allow better recognition of signs. Thus, we chose XGBoost [38] as the final classifier given its efficiency in drawing significant inference, better performance across several iterations, and interpretability. Applying the Hyperparameter Grid Search technique, XGBoost was trained using 250 gradient boosted trees with a maximum tree depth of 7, base score of 0.6 and learning rate of 0.2. Sample of tuning XGBoost models is shown in Table 2.The XGBoost classifier uses the conventional decision trees which are easy to interpret.

Table 2: Hyper parameter tuning of XGBoost model

Number of trees	Depth	Learningrate	Base score	Accuracy (approximate)
200	5	.3	.45	81.45
250	7	0.2	0.6	83.56
300	6	.25	.54	78.32

BERT [41] (Bidirectional Encoder Representations from Transformers) is so designed that it can pretrain deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context at the same time in all layers. Due to this bi-directional mechanism, only the Bert model is tried and implemented with hyperparameter tuning and the selected parameters are shown Table 3. It can befinetuned with just one additional output layer to create state-of-the-art models for a wide range of tasks without much effort on architecture modifications.

Table 3: Hyper parameter selection of BERT masked model

Hyperparameters	**Values**
TRAIN_BATCH_SIZE	32
EVAL_BATCH_SIZE	8
PREDICT_BATCH_SIZE	8
LEARNING_RATE	2e-5
NUM_TRAIN_EPOCHS	3.0
MAX_SEQ_LENGTH	128
WARMUP_PROPORTION	0.1
SAVE_CHECKPOINTS_STEPS	2000
SAVE_SUMMARY_STEPS	600

5. CONCLUSION AND FUTURE SCOPE

The current research work proposed a new novel methodology of combining multiple data set available in the context of Indian sign language recognition to achieve better accuracy. The proposed framework also introduced a new novel scheme of combining a standard sign language detection algorithm with a masked model framework to bridge the gap between isolated sign language recognition and continuous sign language recognition. The proposed scheme has scope of increasing the efficacy of the state of art for Indian sign language recognition manifold by leveraging the data from both the domain(discrete/continuous).

6. REFERENCES

[1] W. H. Organization. (2018) Deafness and hearing loss. [Online]. Available: https://www.who.int/news-room/ fact-sheets/detail/deafness and-hearing-loss.

[2] S. Aly and W. Aly, "Deeparslr: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition," IEEE Access, vol. 8, pp. 83 199–83 212, 2020.

[3] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," Expert Systems with Applications, vol. 150, p. 113336, 2020.

[4] K. Tripathi and N. B. G. Nandi, "Continuous Indian sign language gesture recognition and sentence formation," Procedia Computer Science, vol. 54, pp. 523–531, 2015.

[5] M. Hassan, K. Assaleh, and T. Shanableh, "Multiple proposals for continuous arabic sign language recognition," Sensing and Imaging, vol. 20, no. 1, p. 4, 2019.

[6] S. Albanie, G. Varol, L. Momeni, T. Afouras, J. S. Chung, N. Fox, and A. Zisserman, "Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues," in European Conference on Computer Vision. Springer, 2020, pp. 35–53.

[7] O. Aran, I. Ari, L. Akarun, B. Sankur, A. Benoit, A. Caplier, P. Campr, A. H. Carrillo et al., "Signtutor:An interactive system for sign language tutoring," IEEE MultiMedia, vol. 16, no. 1, pp. 81–93, 2009.

[8] F. S. Botros, A. Phinyomark, and E. J. Scheme, "Electromyography based gesture recognition: Is it time to change focus from the forearm to the wrist?" environments, vol. 14, p. 15, 2020.

[9] R. Fatmi, S. Rashad, and R. Integlia, "Comparing an, svm, and hmm based machine learning methods for American sign language recognition using wearable motion sensors," in 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2019, pp. 0290–0297.

[10] S. Wei, X. Chen, X. Yang, S. Cao, and X. Zhang, "A component-based vocabulary-extensible sign language gesture recognition framework," Sensors, vol. 16, no. 4, p. 556, 2016.

[11] S. Kim, J. Kim, S. Ahn, and Y. Kim, "Finger language recognition based on ensemble artificial neural network learning using armband emg sensors," Technology and Health Care, vol. 26, no. S1, pp. 249–258, 2018.

[12] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," International Journal of Machine Learning and Cybernetics, vol. 10, no. 1, pp. 131–153, 2019.

[13] S. Kim, Y. Ji, and K.-B. Lee, "An effective sign language learning with object detection based roi segmentation," in 2018 Second IEEE International Conference on Robotic Computing (IRC). IEEE, 2018, pp. 330– 333.

[14] M. P. Paulraj, S. Yaacob, M. S. bin Zanar Azalan, and R. Palaniappan, "A phoneme based sign language recognition system using skin color segmentation," in 2010 6th International Colloquium on Signal Processing & its Applications. IEEE, 2010, pp. 1–5.

[15] A. S. Ghotkar and G. K. Kharate, "Vision based real time hand gesture recognition techniques for human computer interaction," Int J Comput Appl, vol. 70, no. 16, pp. 1–8, 2013.

[16] J.-H. Shin, J.-S. Lee, S.-K. Kil, D.-F. Shen, J.-G. Ryu, E.-H. Lee, H.-K. Min, and S.-H. Hong, "Hand region extraction and gesture recognition using entropy analysis," IJCSNS International Journal of Computer Science and Network Security, vol. 6, no. 2A, pp. 216– 222, 2006.

[17] R. Akmeliawati, F. Dadgostar, S. Demidenko, N. Gamage, Y. C. Kuang, C. Messom, M. Ooi, A. Sarrafzadeh, and G. SenGupta, "Towards real-time sign language analysis via markerless gesture tracking," in 2009 IEEE Instrumentation and Measurement Technology Conference. IEEE, 2009, pp. 1200–1204.

[18] X. Wang, M. Xia, H. Cai, Y. Gao, and C. Cattani, "Hidden-markovmodels-based dynamic hand gesture recognition," Mathematical Problems in Engineering, vol. 2012, 2012.

[19] P. Li, T. Zhang, and A. E. Pece, "Visual contour tracking based on particle filters," Image and Vision Computing, vol. 21, no. 1, pp. 111–123, 2003.

[20] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand gesture recognition for sign language using 3dcnn," IEEE Access, vol. 8, pp. 79 491–79 509, 2020.

[21] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 science and information conference. IEEE, 2014, pp. 372–378.

[22] S. Jiang, B. Lv, W. Guo, C. Zhang, H. Wang, X. Sheng, and P. B. Shull, "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via semg and imu sensing," IEEE Transactions on Industrial Informatics, vol. 14, no. 8, pp. 3376–3385, 2018.

[23] Y. Yao and C.-T. Li, "Hand posture recognition using surf with adaptive boosting," in British Machine Vision Conference, 2012.

[24] K. K. Gurjal P, "Real time hand gesture recognition using sift." Int J Electron Electr Eng, vol. 2, no. 3, 2012.

[25] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. Crawford, M. M. Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American sign language recognition using rf sensing," IEEE Sensors Journal, 2021.

[26] S. Jiang, B. Lv, W. Guo, C. Zhang, H. Wang, X. Sheng, and P. B. Shull, "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via semg and imu sensing," IEEE Transactions on Industrial Informatics, vol. 14, no. 8, pp. 3376–3385, 2018.

[27] M. Tariq, A. Iqbal, A. Zahid, Z. Iqbal, and J. Akhtar, "Sign language localization: Learning to eliminate language dialects," in 2012 15th International Multitopic Conference (INMIC). IEEE, 2012, pp. 17–22.

[28] W. Ahmed, K. Chanda, and S. Mitra, "Vision based hand gesture recognition using dynamic time warping for indian sign language," in 2016 International Conference on Information Science (ICIS). IEEE, 2016, pp. 120–125.

[29] N. Kumar, "Sign language recognition for hearing impaired people based on hands symbols classification," in 2017

International Conference on Computing, Communication and Automation (ICCCA). IEEE, 2017, pp. 244–249.

[30] R. A. Kadhim and M. Khamees, “A real-time american sign language recognition system using convolutional neural network for real datasets,” TEM Journal, vol. 9, no. 3, p. 937, 2020.

[31] C. K. Lee, K. K. Ng, C.-H. Chen, H. C. Lau, S. Chung, and T. Tsoi, “American sign language recognition and training method with recurrent neural network,” Expert Systems with Applications, vol. 167, p. 114403, 2021.

[32] O. M. Sincan and H. Y. Keles, “Autsl: A large scale multi-modal turkish sign language dataset and baseline methods,” IEEE Access, vol. 8, pp. 181 340–181 355, 2020.

[33] G. Yuan, X. Liu, Q. Yan, S. Qiao, Z. Wang, and L. Yuan, “Hand gesture recognition using deep feature fusion network based on wearable sensors,” IEEE Sensors Journal, vol. 21, no. 1, pp. 539–547, 2020.

[34] Sridhar, Advaith, Rohith Gandhi Ganesan, Pratyush Kumar, and Mitesh Khapra. "Include: A large scale dataset for indian sign language recognition." In Proceedings of the 28th ACM international conference on multimedia, pp. 1366-1375. 2020.

[35] http://islrtc.nic.in/

[36] https://indiansignlanguage.org/

[37] Cao, Zhe, Tomas Simon, Shih-En Wei, and Yaser Sheikh. "Realtime multi-person 2d pose estimation using part affinity fields." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 72917299. 2017.

[38] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acmsigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.

[39] Meyer, David, Friedrich Leisch, and Kurt Hornik. "The support vector machine under test." Neurocomputing 55, no. 1-2 (2003): 169-186.

[40] Pal, Mahesh. "Random forest classifier for remote sensing classification." International journal of remote sensing 26, no. 1 (2005): 217222.

[41] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805(2018).